

Google Trends -massadata päivittäiselintarvikkeiden tiedonhaun ja ostopäätösten ajoittamisessa sekä myynnin ennustamisessa

Hanne Teräväinen

Tekijä(t) Hanne Teräväinen	
Koulutusohjelma Tietojärjestelmäosaaminen	
Raportin/Opinnäytetyön nimi Google Trends -massadata päivittäiselintarvikkeiden tiedonhaun ja ostopäätösten ajoittamisessa sekä myynnin ennustamisessa	Sivu- ja liitesivumäärä 114 + 76
<p>Vasta viidesosa suomalaisista yrityksistä hyödyntää massadataa liiketoiminnassa, ja harva yritys näkee sen mahdollisuutena kasvattaa liiketoimintaa. Kuitenkin massadataa syntyy koko ajan ympärillämme kiihtyvällä vauhdilla muun muassa Internetin käytöstä. Esimerkiksi hakukoneiden tuottamaa dataa on avoimesti saatavana, mutta yrityksiltä puuttuu osaaamista sen keräämiseen ja jalostamiseen arvoa tuottaviksi analyyseiksi ja palveluiksi.</p> <p>Tämä opinnäytetyö tehtiin osana Business Finlandin rahoittamaa Big Data – Big Business -hanketta, jonka tavoitteena on luoda arvoa yritysten liiketoiminnalle massadatan avulla. Tarkoituksena on kehittää uutta datalähtöistä liiketoimintaa ja parantaa yritysten globaalia kilpailukykyä. Kehittämishanke toteutettiin yhteistyössä kansainvälisen yrityksen kanssa, joka toimii B2B-sektorilla kodin ulkopuolisen mainonnan parissa.</p> <p>Tutkimuksen tavoitteena oli selvittää, mitä yrityksen ulkoista dataa on saatavana kuluttajan ostopäätösprosessin tiedon etsimis- ja ostopäätösvaiheista, miten dataa voi kerätä ja milloin kuluttajat etsivät tietoa ja tekevät ostopäätöksen. Lisäksi tavoitteena oli selvittää, onko Google-hakuintensiteetin ja myynnin välillä riippuvuus, ja voiko niiden perusteella ennustaa myyntimääriä. Tutkittaviksi tapauksiksi valittiin päivittäiselintarvikeryhmistä meijerituotteet, valmisruoat, kasviproteiinit ja sesonkituotteet.</p> <p>Tutkimus toteutettiin vuodenvaihteessa 2019–2020 noudattaen CRISP-DM-prosessia. Data kerättiin Google Trends ja Ubersuggest -hakukoneanalytiikkatyökaluilla ja avoimista myyntitilastoista. Tutkimusmenetelminä käytettiin muun muassa hakutrendi-, korrelaatio-, regressio- ja tilastollista analyysiä. Tutkimuksessa mallinnettiin ostopäätösajankohta viivelukujen avulla ja tehtiin Google Trends -dataa hyödyntävä yksikertainen regressiomalli myyntimäärien ennustamiseen. Analyysit tehtiin IBM SPSS -tilasto-ohjelmalla.</p> <p>Kuluttajien ostopäätösprosessin tiedon etsiminen -vaiheesta on saatavana avointa hakukonedataa lukuisten ilmaisten analytiikkatyökalujen avulla. Sen sijaan ostopäätösvaiheen tietojen hankkiminen on haastavaa, koska B2B-yrityksellä ei ole käytettävissä asiakkaidensa asiakasdataa myynneistä. Mallinnuksessa käytetyn punaviinin Google Trends -hakuindeksien ja myynnin välillä oli voimakas positiivinen korrelaatio Suomessa. Viiveluvuista nähdään, että myynti oli suurinta hakua seuraavan kuukauden aikana. Ennustemalli selitti 49,1 % myynnin vaihtelusta hakuintensiteetin vaihtelulla.</p> <p>Google Trendistä saatavan massadatan ja tutkimuksessa käytettyjen mallien avulla kohdeyritys tunnistaa trendit ja sesongit, kuluttajien kiinnostuksen kohteet ja ajankohdat, ennustaa hakuintensiteetin perusteella myyntimäärät ja ostopäätösajankohdat sekä mittaa altistumista mainonnalle. Näiden avulla yritys voi kehittää toimintaansa tietoon perustuen: optimoida myynnin asiakasyhteydenottojen ajoituksen ja mainonnan ulostulot kuluttajan ostopolun varrella sekä parantaa kilpailukykyä.</p>	
Asiasanat Massadata, big data, Google Trends, ostopäätösprosessi, kuluttajaymmärrys, analyysi, mallinnus, ennustaminen, päivittäiselintarvikkeet	

Sisällys

1	Johdanto	1
1.1	Big Data – Big Business -hankkeella arvoa yritysten liiketoiminnalle massadataa hyödyntämällä	2
1.2	Toimeksiantajan ja kohdeyrityksen esittely	2
1.3	Tavoitteet ja tutkimuskysymykset	3
1.4	Rajaukset	4
1.5	Tutkimuksen lähestymistavat	4
1.6	CRISP-DM-prosessi	5
1.6.1	Liiketoiminnan ymmärtäminen	6
1.6.2	Datan ymmärtäminen	7
1.6.3	Datan valmistelu	7
1.6.4	Mallinnus	7
1.6.5	Arviointi	7
1.6.6	Käyttöönotto	7
1.7	Raportin rakenne	8
1.8	Keskeiset käsitteet	8
2	Kuluttajan ostopäätösprosessi	9
2.1	Kotlerin ja Kellerin kuluttajan ostopäätösprosessin viisivaiheinen malli	9
2.1.1	Tarpeen tunnistaminen	10
2.1.2	Tiedon etsiminen	11
2.1.3	Vaihtoehtojen vertailu	12
2.1.4	Ostopäätös	12
2.1.5	Oston jälkeinen käyttäytyminen	13
2.2	Kuluttajan omni-channel-ostoprosessi	14
3	Ulkomainonta liiketoimintana	15
3.1	Ulkomainonta massamediavälineenä	15
3.2	Ulkomainonnan vaikutus kuluttajan ostopäätösprosessiin	16
4	Tutkimusmenetelmien ja työkalujen kuvaus sekä teoreettinen tausta	18
4.1	Puolistrukturoitu haastattelu	18
4.2	Dokumenttianalyysi	18
4.3	Elintarvikenimikkeiden määrittäminen COICOP-HBS-luokituksella	18
4.4	Hakutrendien analysointi	19
4.5	Kahden muuttujan välisen riippuvuuden tutkiminen	20
4.6	Ristikorrelaatio- ja spatiotemporaalinen analyysi	21
4.7	Pearsonin korrelaatioanalyysi	22
4.8	Regressioanalyysi	23
4.9	Tilastollinen analyysi	27

4.10	Google Trends -hakukoneanalytiikkatyökalu	30
4.11	Ubersuggest-hakukoneanalytiikkatyökalu	32
4.12	Excel-taulukkolaskentaohjelma	33
4.13	IBM SPSS Statistics -tilasto-ohjelma	33
4.14	Tilastokeskuksen Tee oma karttaesitys -palvelu	33
5	Tutkimuksen toteutus	34
5.1	Liiketoiminta- ja kuluttajaymmärryksen parantaminen	34
5.2	Datan ymmärtäminen, kerääminen ja valmistelu	35
5.2.1	Päivittäiselintarvikkeiden hakuajankohdat	35
5.2.2	Punaviinin Google-hakujen ja ostamisen välinen viive	36
5.2.3	Punaviinin myynnin ennustemalli	37
5.3	Analyytit ja mallinnus.....	38
5.3.1	Päivittäiselintarvikkeiden hakuajankohdat	38
5.3.2	Punaviinin Google-hakujen ja ostamisen välinen viive	43
5.3.3	Punaviinin myynnin ennustemalli	43
6	Tutkimuksen tulokset.....	46
6.1	Kuluttajaymmärrystä data-analytiikalla	46
6.2	Päivittäiselintarvikkeiden hakuajankohdat	58
6.2.1	Meijerituotteet/voi	58
6.2.2	Meijerituotteet/jäätelö	59
6.2.3	Meijerituotteet/juusto	60
6.2.4	Valmisruoat/maksalaatikko.....	61
6.2.5	Valmisruoat/kinkkukiusaus	62
6.2.6	Valmisruoat/kasvissosekeitto	62
6.2.7	Kasviproteiinit/Nyhtökaura.....	63
6.2.8	Kasviproteiinit/soijarouhe	64
6.2.9	Kasvisproteiinit/Härkis.....	65
6.2.10	Sesonkituotteet/porkkanalaatikko.....	66
6.2.11	Sesonkituotteet/lanttulaatikko.....	67
6.2.12	Sesonkituotteet/joulukinkku	69
6.2.13	Punaviini	70
6.3	Punaviinin Google-hakujen ja ostamisen välinen viive	71
6.4	Punaviinin myynnin ennustemalli	73
7	Pohdinta.....	82
7.1	Tulosten tarkastelu.....	82
7.2	Tutkimuksen luotettavuus	84
7.3	Eettiset näkökohdat	87
7.4	Johtopäätökset sekä kehittämis- ja jatkotutkimusehdotukset.....	87
7.5	Oman oppimisen arviointi.....	98

Lähteet	100
Liitteet.....	115
Liite 1. Keskeiset käsitteet, lyhenteet ja englanninkieliset termit	115
Liite 2. Sesonkituotteiden kartoitus: hakumäärät ja -piikit	123
Liite 3. Datojen mallitaulukot.....	124
Liite 4. Ohjelmien ja palveluiden asetukset.....	126
Liite 5. Elintarvikeryhmien tuotenimikkeiden sekä hakusanojen ja -volyymien määrittäminen	130
Liite 6. Ubersuggest-työkalun tuottamat hakuvolyymit ja hakusanaideat	135
Liite 7. Tuotenimikkeiden hakijoiden ikäryhmät.....	136
Liite 8. Maakuntien ristikorraatiotaulukot.....	137
Liite 9. Maakuntien ristikorraatiokertoimet viiveluvuilla 0, 1 ja 2 ja tilastollinen merkitsevyys	143
Liite 10. Maakuntien ristikorraatiokertoimien luokittelu karttojen tekemistä varten...	144
Liite 11. Pearsonin korraatiomatriisi	145
Liite 12. Voin hakuajankohdat	146
Liite 13. Jäätelön hakuajankohdat	150
Liite 14. Juuston hakuajankohdat	154
Liite 15. Maksalaatikon hakuajankohdat	158
Liite 16. Nyhtökauran hakuajankohdat	161
Liite 17. Soijarouheen hakuajankohdat.....	165
Liite 18. Härkiksen hakuajankohdat.....	168
Liite 19. Porkkanalaatikon hakuajankohdat	172
Liite 20. Lanttulaatikon hakuajankohdat	176
Liite 21. Joulukinkun hakuajankohdat.....	180
Liite 22. Punaviinin hakuajankohdat	184
Salaiset liitteet	188
Salainen liite 1. Kehittämisprojektin suunnitelma	188
Salainen liite 2. Kohdeyrityksen kehitysjohtajan Anon haastattelulomake.....	189
Salainen liite 3. Kohdeyrityksen kehitysjohtajan Anon litteroitu haastattelu.....	190

1 Johdanto

Google. Maailman arvokkain yritys tuottaa Internet-palveluita (Lampela 2018). Sen haku-koneesta etsitään globaalisti tietoa lähes kuusi miljardia kertaa päivässä (Google 2020a). Valtavia määriä massadataa syntyy koko ajan kiihtyvällä vauhdilla muun muassa Internetin käytöstä, sosiaalisesta mediasta sekä yritysten omista älylaitteista ja sensoreista lähes reaaliajassa. Data itsessään ei ole merkityksellistä, vaan keräämisen ja jalostuksen tulok-sena syntyvät arvoa tuottavat analyysit, palvelut ja tuotteet. (Ali-Yrkkö, Mattila, Pajarinen & Seppälä 2019, 9, 10.) Maailmanlaajuisesti massadatateknologian markkina-arvo oli 36,8 miljardia dollaria vuonna 2018 ja sen ennustetaan kasvavan vuoteen 2026 mennessä 104,3 miljardiin dollariin. Useiden tutkimusten mukaan massadata-analytiikkatyökalujen hyödyntäminen on parantanut tuottavuutta 5–10 %. (Fortune Business Insights 2020.)

Vasta vajaa viidesosa suomalaisista vähintään 10 työntekijän yrityksistä hyödynsi liiketoi-minnassaan massadataa vuonna 2018. Vertailussa muihin EU-maihin Suomi sijoittui nel-jänneksi. Massadataa hyödynnettiin yleisemmin sitä enemmän mitä suurempi yritys oli ky-seessä. Kuitenkin Suomessa suurten ja keskisuurten yritysten massadatan hyödyntämi-sen vauhti on hidastunut. (Ali-Yrkkö ym. 2019, 9–10, 17–18.) Suomessa eri toimialoista massadataa hyödynsivät eniten informaatio- ja viestintäpalvelut (42 %) sekä ICT-alat (42 %), jotka sijoittuvat EU-vertailussa jaetulle toiselle sijalle. Kauppa (15 %) sijoittui toimiala-vertailussa Suomessa kuudennelle ja EU:ssa viidennelle sijalle. (Ali-Yrkkö ym. 2019, 20).

Tutkimuksen mukaan suomalaiset yritykset (n = 1 189) hyödynsivät massadataa eniten päätöksenteossa, asiakas- ja markkina-analyysissä sekä uustuotekehityksessä. Yritykset kokivat, että massadatan arvontuotto näkyy liiketoiminnassa tietoon perustuvissa päätök-sissä sekä kilpailuaseman, tuotteiden laadun ja asiakaskokemuksen paranemisessa. Mer-kittävä este yrityksissä massadatan hyödyntämiselle oli datan käsittelyn ja analysoinnin osaamisvaje. (Antikainen, Eskelinen, Koski, Niemi, Pajarinen, Pyykkönen & de Vries 2016, 7.) Harva yritys osaa vielä nähdä massadatan hyödyntämisen potentiaalisena mah-dollisuutena liikevaihdon kasvattamiseen (Ali-Yrkkö ym. 2019, 12; Antikainen ym. 2016, 7).

Tämä opinnäytetyö on tehty osana Big Data – Big Business -hanketta, joka osaltaan ha-luaa vastata edellä kuvattuun haasteeseen, sillä hankkeen tavoitteena on luoda arvoa yri-tysten liiketoiminnalle massadataa hyödyntämällä. Hanke on esitelty tarkemmin kohdassa 1.1. Haaga-Helia ammattikorkeakoulun toimeksiannosta toteutettiin kohdeyritykselle useita kehittämisprojekteja. Tämä kehittämisprojekti on osaprojektiin ”Data-analytiikan työ-

kaluilla kerättävät tiedot ja mahdollisuudet” liittyvä itsenäinen kokonaisuus, joka on toteutettu yhteistyössä kohdeyrityksen kanssa. Toimeksiantaja ja yritys on esitelty kohdassa 1.2. Tutkimuksen tavoitteet, tutkimuskysymykset, rajaukset, tutkimuksen lähestymistavat ja prosessi on kuvattu kohdissa 1.3.–1.6. Raportin rakenne kokonaisuudessaan on esitelty kohdassa 1.7. ja keskeiset käsitteet kohdassa 1.8.

1.1 Big Data – Big Business -hankkeella arvoa yritysten liiketoiminnalle massadataa hyödyntämällä

Big Data – Big Business on Business Finlandin rahoittama hanke, jonka tavoitteena on big dataa eli massadataa hyödyntämällä luoda uutta akateemista ja käytännönläheistä tietoa uuden liiketoiminnan synnyttämiseksi. Lisäksi tavoitteena on edistää suomalaisten B2B-yritysten massadataan liittyvää osaamista ja kykyä tunnistaa uusia liiketoimintamahdollisuuksia kansainvälisen kilpailukyvyn edistämiseksi. Hanke yhdistää massadatan hyödyntämisen datakeskeiseen ja liiketoiminnalliseen näkökulmaan. Sen tuloksia on tarkoitus hyödyntää massadataan perustuvan liiketoiminnan kehittämisessä ja palveluiden laajentamisessa globaaleille markkinoille. Hankkeella on useita kumppaniyrityksiä, ja tutkimusorganisaatiot ovat Haaga-Helia ammattikorkeakoulu, VTT ja Itä-Suomen yliopisto. (Haaga-Helia 2019a.)

1.2 Toimeksiantajan ja kohdeyrityksen esittely

Tämän kehittämisprojektin toimeksiantaja on Haaga-Helia ammattikorkeakoulu, jonka ydinosaamista ovat palveluliiketoiminnan osaamispalvelut, tutkimus ja kehittäminen (Haaga-Helia 2019b).

Kohdeyritys on kansainvälinen yritys, jonka toimiala on kodin ulkopuolinen mainonta eli ulkomainonta. Yrityksen liikevaihto oli noin 30 miljoonaa euroa vuonna 2018, ja se työllistää useita kymmeniä ammattilaisia (Finder 2020). Tällä hetkellä kohdeyritys kerää dataa useista eri lähteistä muun muassa tutkimuksista, ulkomainospintojen havainnointitutkimuksista ja erilaisista julkisista lähteistä. Kuluttajien profiloimiseksi hyödynnetään kyselytutkimuksen lisäksi monipuolisesti erilaista saatavilla olevaa avointa dataa. Tämän datayhdistelmän avulla yritys pystyy tarjoamaan mainostajille mitattavia mainosratkaisuja, joiden tavoitteena on saavuttaa mainostajan kohderyhmät entistä tehokkaammin. (Anon 2019.)

1.3 Tavoitteet ja tutkimuskysymykset

Tämän tutkimuksen tärkeimpänä tavoitteena on tuottaa kohdeyritykselle uudenlaista analysoitua ja konkreettista tietoa kuluttajien ostopäätösprosessin tiedon etsimis- ja ostopäätösvaiheiden ajankohdista liiketoiminnan kehittämiseksi ja kuluttajaymmärryksen parantamiseksi. Tavoitteena on selvittää, milloin kannattaa olla läsnä kuluttajan ostopolulla tiedon etsimis- ja ostopäätösvaiheessa, jotta voidaan vaikuttaa kuluttajan ostopäätökseen. Tutkimuksessa tarkastellaan FMCG-vertikaalin neljää elintarviketuoteryhmää: meijerituotteita, valmisruokia, kasviproteiineja ja sesonkituotteita. Lisäksi tavoitteena on kartoittaa, mitä kuluttajaymmärrystä lisäävää dataa on saatavana ostoprosessin tiedon etsimisen ja ostopäätöksen vaiheista, ja mitä dataa data-analytiikan työkaluilla voi kerätä. Tutkimuksen tavoitteena on lisäksi luoda ja kuvata analyysimallit, jotka kehittävät kohdeyrityksen datan hyödyntämisen toimintamalleja esimerkiksi myyntimäärän ennustamisessa.

Tutkimusongelmaa lähestyttiin monialaisesti yhdistämällä hakukone- ja data-analytiikkaa uudella tavalla, sillä kohdeyritys odottaa hyötyvänsä tutkimuksesta saamalla tietoa kuluttajien kiinnostuksen kohteiden ajoituksista. Pitkän aikavälin hakujen tuntemus auttaa myyntiä ottamaan yhteyttä mainostajiin otollisena hetkenä. Lyhyen aikavälin hakujen avulla mainonta voidaan kohdistaa lähelle ostopäätöstä, ja aktivoida kuluttaja ostamaan haluttu tuote. Jos Google Trendsin hakuindeksien perusteella voidaan ennustaa tuleva myynti, on se selkeä kilpailuetu.

Tutkimuksen pääongelma on muotoiltu pääkysymykseksi:

Milloin kuluttajat hakevat tietoa netistä päivittäiselintarvikkeista ja tekevät ostopäätökset, ja mitä dataa data-analytiikalla voi kerätä ja miten hyödyntää?

Pääkysymys on eritelty tarkentaviin kysymyksiin alaongelmien perusteella:

TK1: Mitä kuluttajaymmärrystä lisäävää dataa on saatavana ostoprosessin tiedon etsimisen ja ostopäätöksen vaiheista, ja mitä dataa data-analytiikan työkaluilla voi kerätä (huomioiden tiedon yksityisyyteen liittyvän lainsäädännön (GDPR))?

TK2: Milloin kuluttajat hakevat tietoa Googlestä valituista päivittäiselintarvikkeista? Miten trendit, sesongit, kuukaudet, viikot, viikonpäivät, kellonajat ja sijainnit vaikuttavat eri päivittäiselintarvikkeiden Google-hakuihin? Miten kuluttajaryhmät eroavat tiedonhakemisessa Googlestä eri päivittäiselintarvikkeiden kohdalla?

TK3: Kuinka pitkä viive on punaviinin Google-hakujen ja ostamisen välillä Suomessa ja eri maakunnissa?

TK4: Onko punaviinin Google-hakuintensiteetillä ja myynnillä positiivinen riippuvuus Suomessa, ja voiko niiden perusteella ennustaa myyntimäärää?

Peittomatriisiin (Haaga-Helia ammattikorkeakoulu 2020) (taulukko 1) on koottu luvun numeroittain tutkimusongelmiin liittyvä teoreettinen viitekehys, tulokset ja aineistot.

Taulukko 1. Peittomatriisi (Haaga-Helia ammattikorkeakoulu 2020)

Tutkimusongelma	Viitekehys	Tulokset	Aineistot
	Luvun numero		
Alaongelma 1	2, 3, 4.1., 4.2., 4.10., 4.11.	6.1.	Dokumentit, haastattelu
Alaongelma 2	2, 3, 4.1.–4.4., 4.9.–4.13.	6.2.	Edellisten lisäksi indeksi- ja myyntidata
Alaongelma 3	2, 3, 4.1., 4.2., 4.6., 4.9., 4.10., 4.12.–4.14.	6.3.	
Alaongelma 4	2, 3, 4.1., 4.2., 4.5., 4.7.–4.10., 4.12., 4.13.	6.4.	

1.4 Rajaukset

Tutkimus keskittyy yrityksen ulkoiseen dataan, joka on kerättävissä data-analytiikan työkaluilla. Yrityksen sisäisestä datasta kerrotaan lyhyesti, koska halutaan antaa kokonaiskuva saatavilla olevasta datasta ja tuoda esiin sisäisen ja ulkoisen datan erot. Sisäisen datan tarkempi tarkastelu rajattiin pois, koska B2B-yritys tarvitsee tietoa asiakkaidensa asiakkaista eli kuluttajista, jolloin esimerkiksi omaa transaktiodataa ei ole käytettävissä tähän tarkoitukseen.

Tutkimuksen empiirisestä osasta rajattiin pois kuluttajan ostopäätösprosessin vaiheista tarpeen tunnistaminen, vaihtoehtojen vertailu ja oston jälkeinen käyttäytyminen. Empiirinen osa toteutettiin avoimella datalla, ja datat kerättiin maksuttomilla analytiikkatyökaluilla. Vaihtoehtoista rajattiin pois maksulliset datat ja keräämisen työkalut.

Tutkimuksen laajuutta rajattiin valitsemalla FMCG-vertikaalista neljä tuoteryhmää: meijerituotteet, valmisruoat, kasviproteiinit ja sesonkituotteet. Kustakin ryhmästä valittiin kolme tuotenimikettä tarkasteluun. Valmisruokatuoteryhmä on hyvin laaja, joten määritettiin tutkimus koskemaan ainoastaan valmisaterioista laadit, keitot, salaattit ja pasta-ateriat -alaryhmää. Ennustemallissa ei tehty maakuntakohtaisia tarkasteluja työn mittavuuden takia. Myös tutkimustietoon perustuvat argumentit ulkomainonnan vaikuttavuudesta rajattiin tutkimuksen ulkopuolelle.

1.5 Tutkimuksen lähestymistavat

Opinnäytetyöllä on yrityksen toiminnan kehittämistä palveleva päämäärä, jolloin sen avulla tuotetaan tietoa tutkittavasta ilmiöstä päätöksentekoa varten. Tutkimusta varten hankittiin

taustatietoja toimeksiantajayrityksestä kvalitatiivisin eli laadullisin menetelmin käyttäen puolistrukturoitua haastattelua ja myös ilmiön ymmärtämiseksi dokumentteja.

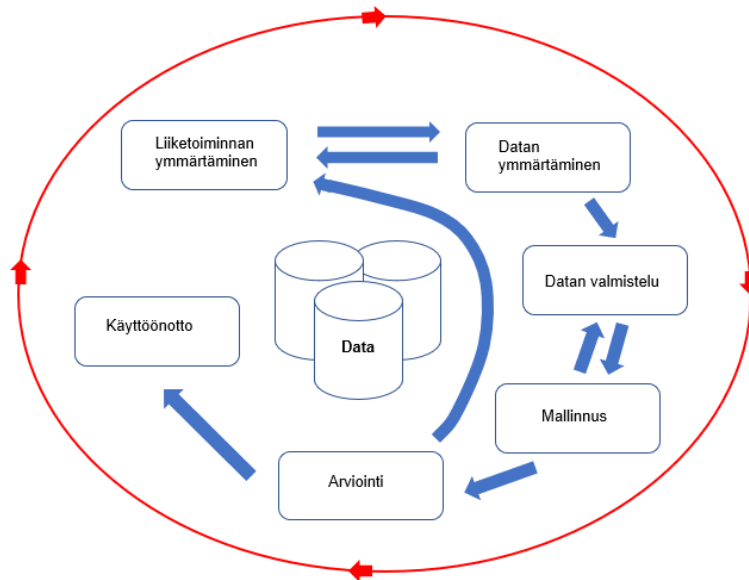
Tapaustutkimuksella pyritään muodostamaan monipuolinen ja syvälle menevä kuva tutkitavasta ilmiöstä, jolloin aineistonkeruu- ja analyysimenetelminä käytetään laadullisia ja määrällisiä menetelmiä (Kananen & Makkonen (toim.) 2015, 76). Kohdeyritys valitsi tämän tutkimukseen caseiksi FMCG-vertikaalin ja tutkittaviksi tapauksiksi neljä elintarviketuoter ryhmää: meijerituotteet, valmisruoat, kasviproteiinit ja sesonkituotteet. FMCG-tuotteilla tarkoitetaan päivittäistavaroita eli nopeasti liikkuvia, kuluttajille suunnattuja kulutustavaroita, jotka ostetaan usein ruokaostojen yhteydessä (Wikipedia 2020a). Päivittäistavaroita ovat elintarvikkeet ja kulutustuotteista hygieniatuotteet, teknokemian- ja tupakkatuotteet sekä lehdet ja päivittäiskosmetiikka, joita nimensä mukaisesti käytetään päivittäin. FMCG-vertikaalin merkitys päivittäistavaramyynnistä on suuri, noin 80 %, joka vastaa lähes 15 miljardia euroa vuodessa (Päivittäistavarakauppa ry 2020.) Elintarvikkeiden kuluttajabrändien valmistajat ja markkinoijat ovat kohdeyritykselle tärkeitä asiakkaita, ja myyjien oikea-aikaisella kontaktoinnilla voidaan tukea asiakkaan päätöksentekoa. Ulkomainonnalla voidaan puolestaan edistää brändin myyntiä vaikuttamalla kuluttajan ostoprosessin eri vaiheisiin ja saattamaan osto päätökseen.

Tutkimuksen analyttiseksi päälähestymistavaksi valittiin kvantitatiivinen eli määrällinen ote, koska tutkimuksessa on tarkoitus mallintaa muun muassa ennustamalla tutkittavia asioita, joka edellyttää yleistämistä ja tilastollisten menetelmien käyttämistä analysoinnissa. Yleistämiseen pyrkivän määrällisen tutkimuksen taustalla tulee olla vankka tuntemus ilmiöstä, joka edellyttää aiheeseen liittyvien teorioiden tuntemusta. Määrällisessä tutkimuksessa aineisto kerätään yleensä kyselyillä (Kananen & Makkonen (toim.) 2015, 73–74.), mutta tässä tutkimuksessa käytetään ulkoisen datan analysointimallin rakentamista varten valmista kerättyä tilastoaineistoa ja Google Trends -palvelusta itse kerättyjä dataja. Määrällisen tutkimuksen tulokset esitetään erilaisina lukuina mukaan lukien tilastolliset tunnusluvut, ja laadullisen tutkimuksen tulokset ovat tekstimuodossa (Kananen & Makkonen (toim.) 2015, 66, 75, 84). Tutkimuksen pohjana on käytetty lukuisia teorioita, tutkimuksia ja ajankohtaista ammattitietoa.

1.6 CRISP-DM-prosessi

Opinnäytetyö toteutettiin analytiikkahankkeisiin soveltuvaa CRISP-DM-prosessia eli tiedonlouhinnan poikkialaista standardiprosessia soveltaen (kuvio 1). Kuusivaiheinen prosessi käynnistettiin liiketoimintaymmärryksen luomisella, jonka jälkeen siirryttiin datan ym-

märtämiseen, valmisteluun ja mallintamiseen. Mallien arvioinnin jälkeen siirryttiin käyttöönottoon. Prosessi eteni eri vaiheiden välillä iteroiden, ja välillä prosessissa palattiin taaksepäin aikaisempiin vaiheisiin. CRISP-DM-prosessin kuvauksessa on käytetty lähteenä IBM:n verkkosivuja. (IBM 2012.)



Kuvio 1. Tiedonlouhinnan poikkialainen standardiprosessi (CRISP-DM) (mukaillen IBM 2012)

1.6.1 Liiketoiminnan ymmärtäminen

Kehittämiprojektin aluksi määritettiin kohdeyrityksen liiketoiminnalliset tavoitteet opinnäytetyölle, jotka on kuvattu kohdassa 1.3. Perehdyttiin yrityksen toimintaan ja toimialaan dokumenttiaineistoista, henkilökohtaisilla keskusteluilla toimeksiantajan ja kohdeyrityksen kanssa sekä kohdeyrityksen kehitysohjattajan haastattelulla.

Saatiin selville, että kohdeyritykselle on tärkeää tuntea tuotteiden hakuajankohdat sekä pitkällä että lyhyellä aikajänteellä. Pitkän aikavälin hakujen tuntemus auttaa suunnittelemaan mainontaa vuositasolla ja kontaktoimaan potentiaaliset mainostaja-asiakkaat oikea-aikaisesti. Lyhyen aikavälin hauilla saadaan selville, milloin kuluttajat ajattelevat tiettyjä tuotteita. Näin tuotteiden mainonta saadaan kohdistettua lähelle ostopäätöstä, ja voidaan vaikuttaa siihen, että kuluttaja valitsee tietyn brändin ostoskoriinsa. (Anon 8.10.2019.)

Data ohjaa kohdeyrityksen liiketoimintaa ja yhteistyötä eri toimijoiden kesken. Yritys onkin kiinnostunut saamaan tietoa, millaisia uusia datalähteitä olisi hyödynnettävissä liiketoiminnassa. Yrityksen liiketoiminnan ydin ovat kontaktit, jolloin tärkeää tietoa ovat kuluttajatyypit, ostoaiheet eri vuodenaikoina, määrät ja sijainnit. Kuluttajien kiinnostuksen kohteet

vaihtelevat ajan mukaan, jolloin ajoitus on tärkeä tekijä mainonnan onnistumisen kannalta. (Anon 8.10.2019.)

Ratkaistavaksi ongelmaksi tunnistettiin kuluttajien valikoitujen päivittäiselintarvikkeiden tiedonhaun ja ostopäätöksen ajoittaminen sekä myynnin ennustaminen massadataa hyödyntäen data-analytiikalla. Seuraavaksi mietittiin tiedonlouhinnan tavoitteet, mahdolliset rajoitteet ja tehtiin kehittämisprojektin suunnitelma (salainen liite 1).

1.6.2 Datan ymmärtäminen

Selvitettiin, millaista dataa tarvitaan ja millaista on saatavana, mitkä ovat datalähteet ja millä työkaluilla datat voi kerätä. Datat kerättiin ja kuvailtiin sekä varmistettiin datojen käytökelpoisuus ja laatu. Datan laatuun vaikuttavat muun muassa puutteet, virheet ja koodauksen epäjohdonmukaisuudet. Datan ymmärtämiseen liittyvät työvaiheet on kuvattu kohdassa 5.2.

1.6.3 Datan valmistelu

Tehtiin päätös käytettävistä datoista. Datat valmisteltiin, tarvittaessa puhdistettiin, yhdisteltiin eri datasettejä ja järjestettiin käyttökelpoiseen muotoon. Datan valmistelun eri vaiheet on kuvattu kohdassa 5.2.

1.6.4 Mallinnus

Selvitettiin kuluttajien tekeminen päivittäiselintarvikkeiden hakuajankohdat, ja mallinnettiin punaviinin tiedon hakemisen ja oston välinen viive sekä myynnin ennustemalli käyttäen avuksi Google Trends -dataa. Mallinnustekniikat ja toteuttaminen on kuvattu kohdassa 5.3.

1.6.5 Arviointi

Arvioitiin projektin ja prosessin onnistumista vertaamalla saavutettuja tuloksia asetettuihin tavoitteisiin sekä pohdittiin jatkotoimenpiteiden tarvetta. Arviointivaihe on kuvattu luvussa 7.

1.6.6 Käyttöönotto

Tehtiin loppupäätelmät mallin hyödyntämisestä kohdeyrityksen liiketoiminnassa, kuvattiin raporttiin projekti vaiheittain, esiteltiin suullisesti projektin tulokset kohdeyritykselle työn keski- ja loppuvaiheessa sekä tehtiin PowerPoint-esitys yrityksen sisäistä käyttöä varten.

Opinnäytetyön ohjaaja ja kirjoittaja laativat lisäksi artikkelin tutkimuksen löydöksistä julkaistavaksi Haaga-Helia ammattikorkeakoulun verkkolehdessä eSignalsissa.

1.7 Raportin rakenne

Opinnäytetyön ensimmäinen luku on tutkimuksen johdanto-osuus. Siinä kuvataan Big Data – Big Business -hanke, toimeksiantaja ja kohdeyritys, tutkimuksen tavoitteet ja tutkimuskysymykset, rajaukset, tutkimuksen lähestymistavat ja prosessi. Lukuihin 2–4 on koottu tutkimuksen tietoperusta. Luku 2 käsittelee kuluttajan ostopäätösprosessia, ja luku 3 taustoittaa ulkomainontaa liiketoimintana. Luvussa 4 kuvataan tutkimusmenetelmät ja työkalut teoreettisine taustoineen, joten tietoperustan ohella se kuuluu myös empiiriseen osaan. Tutkimuksen empiirinen osuus muodostuu luvuista 4–6, joissa kuvataan tutkimusmenetelmät ja työkalut, tutkimuksen toteutus ja tulokset. Pohdintaa käsittelevä luku 7 vastaa tutkimuskysymyksiin ja tarkastelee tuloksia, arvioi tutkimuksen luotettavuutta, puntaroi eettisiä näkökohtia, vetää yhteen johtopäätökset sekä antaa kehittämis- ja jatkotutkimusehdotuksia. Lopuksi arvioidaan tekijän omaa oppimista opinnäytetyöprosessissa.

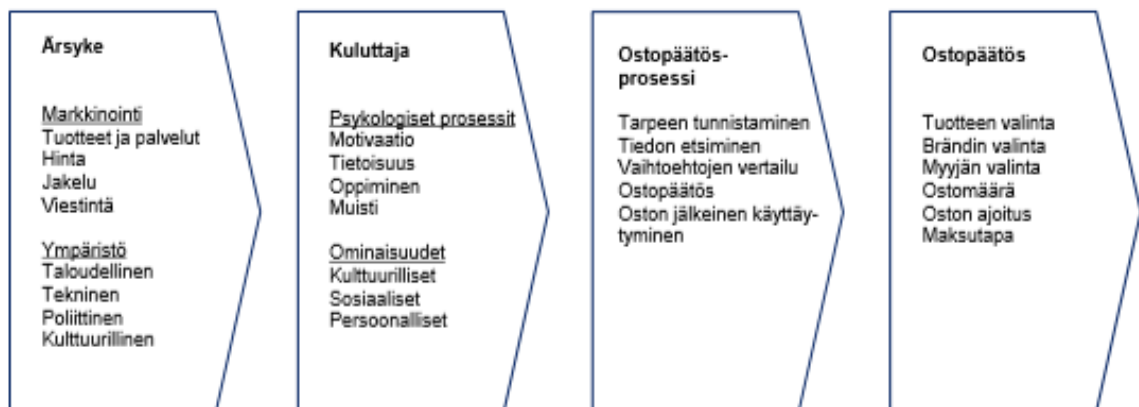
1.8 Keskeiset käsitteet

Keskeiset käsitteet, lyhenteet ja pääasiassa SPSS-ohjelman taulukoiden lukemista helpottavat englanninkieliset termit on koottu liitteeseen 1.

2 Kuluttajan ostopäätösprosessi

Kuluttajan matkaa ostavaksi asiakkaaksi kutsutaan ostopäätösprosessin ohella myös ostoprosessiksi, ostopoluksi, ostosuppiloksi ja myyntitunneliksi (Kivelä 24.6.2019). Asiakkaan ostopolulla tarkoitetaan suunniteltua tai suunnittelematonta tapahtumasarjaa, jossa asiakkaat saavat tietoa, ostavat ja ovat vuorovaikutuksessa yrityksen tarjoamien hyödykkeiden, tavaroiden, palveluiden tai elämysten kanssa (Norton & Pine II 2013). Kotler & Keller (2016, 195) ovat kuvanneet kuluttajan ostopäätösprosessin viisivaiheisena mallina. Kuluttajalla tarkoitetaan henkilöä, joka käyttää hyödykkeitä (Raijas 25.2.2016), kun asiakkaalla puolestaan tarkoitetaan henkilöä, joka ostaa hyödykkeen suorittaen maksun sopimuksen mukaan. Yrityksen ulkopuolinen asiakassuhde voi olla joko suora tai epäsuora. (Wikipedia 2020b.)

Ärsyke–vaste-malli auttaa ymmärtämään kuluttajan käyttäytymistä (kuvio 2). Kun markkinointi ja ympäristön ärsykkeet saavuttavat kuluttajan tietoisuuden, ne käynnistävät psykologiset prosessit, jotka yhdistyvät kuluttajan henkilökohtaisiin ominaisuuksiin. Tämän tuloksena on ostopäätösprosessi, jota seuraa ostopäätös. (Kotler & Keller 2016, 187.)



Kuvio 2. Kuluttajakäyttäytymisen ärsyke–vaste-malli (mukaillen Kotler & Keller 2016, 187)

2.1 Kotlerin ja Kellerin kuluttajan ostopäätösprosessin viisivaiheinen malli

Kuluttajan ostopäätökseen vaikuttaminen edellyttää taustalla toimivan psykologisen prosessin ymmärtämistä, jonka avainkysymyksiä ovat kuka, mitä, milloin, missä, miten ja miksi (Kotler & Keller 2016, 194). Motivaatio, tietoisuus, oppiminen ja muisti ovat neljä psykologista pääprosessia, jotka vaikuttavat kuluttajakäyttäytymiseen. Jotta päästäisiin kiinni, miten kuluttaja tekee ostopäätöksen, tulee tunnistaa vaikuttavat roolit: perehdyttäjät/herättäjät, vaikuttajat, päättäjät, ostajat ja käyttäjät. (Kotler & Keller 2016, 205.)

Kotlerin ja Kellerin (2016, 205) mukaan kuluttajia voi palvella tehokkaammin, kun tuntee kuluttajan käyttäytymiseen vaikuttavat tekijät:

- Kulttuuriset (kulttuuri, alakulttuuri ja sosiaaliluokka)
- Sosiaaliset (viiteryhmät, perhe, sosiaaliset roolit ja statukset)
- Persoonalliset (ikä, elämänvaihe, ammatti, taloudellinen tilanne, elämäntyyli, persoonallisuus ja itsetuntemus).

Kuluttajan demografiset, psykologiset ja sosiaaliset ominaisuudet vaikuttavat suuresti ostopäätöksiin, ja vaikutusta on myös yksittäisessä ostopäätöksessä ostajan aktiivisuudella, tuotteiden erilaisuudella ja ostajan sitoutuneisuudella (Bergström & Leppänen 2015, 120).

Kuluttajan ostoprosessi (kuvio 3) alkaa kauan ennen itse ostotapahtumaa, ja sillä on pitkäaikaisia vaikutuksia. Kuluttajat eivät aina etene järjestelmällisesti prosessin viiden vaiheen mukaan, vaan he voivat ohittaa vaihteita tai edetä päinvastaisessa järjestyksessä. He voivat edetä myös vaihtoehtoisesti netissä tai kivijalassa. Esimerkiksi kuluttajan vakituisesti käyttämän päivittäistavaran osto etenee usein tarpeesta suoraan ostopäätökseen. (Kotler & Keller 2016, 195.)



Kuvio 3. Kuluttajan ostopäätösprosessin viiden kohdan malli (mukaillen Kotler & Keller 2016, 195)

Päivittäistavaroiden ostaminen on usein rutiininomaista ja perustuu tottumukseen. Kuluttaja tekee rutiiniosastoja, jos hänellä on aikaisempaa kokemusta tuotteista, tuotteiden ostaminen on säännöllistä, ostoriski on pieni ja tuotteisiin käytettävä rahasumma on alhainen. Kuluttaja ostaa samoja, tuttuja tuotteita säästääkseen aikaa ja vaivaa. Kuluttajan ostopäätösprosessi vaihtelee henkilöstä, tuotteesta ja tilanteesta johtuen. Rutiiniosastoissa prosessin vaiheet voivat lyhentyä tai jäädä väliin ja ostopäätös syntyy monesti vasta myymälässä. (Bergström & Leppänen 2015, 121.)

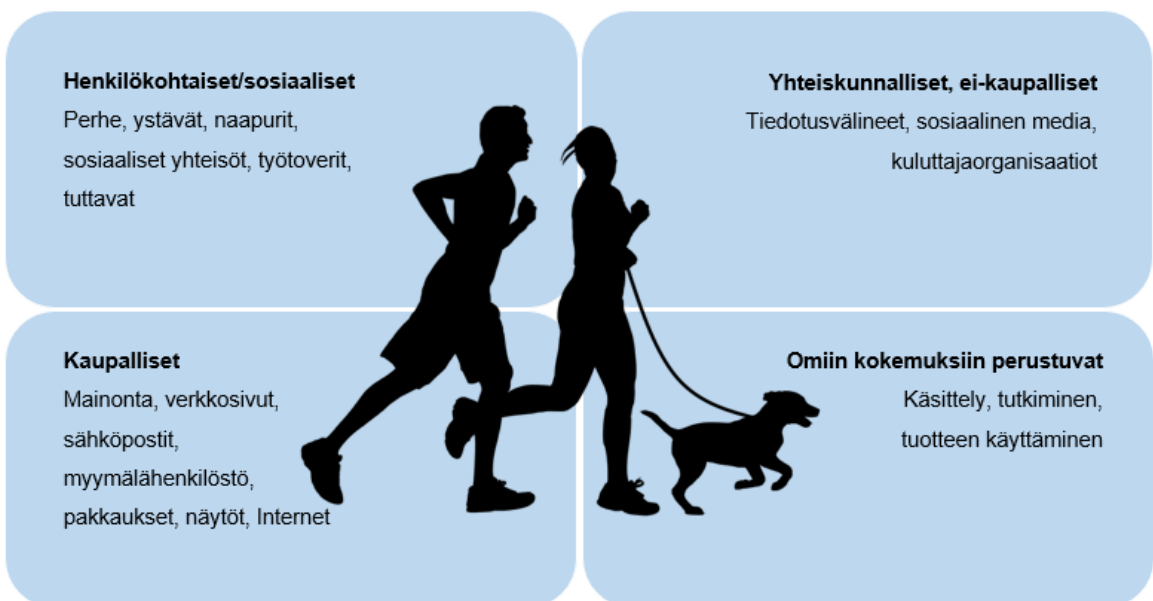
2.1.1 Tarpeen tunnistaminen

Ostoprosessi alkaa, kun kuluttaja havaitsee ratkaistavan ongelman tai jokin ärsyke herättää tarpeen (Kotler & Keller 2016, 195) esimerkiksi jokin tuote on loppunut tai rikkoutunut (Bergström & Leppänen 2015, 121). Ärsyke tarkoittaa herätettä, jonka seurauksena kuluttaja havaitsee tarpeen ja motivoituu toimimaan. Se voi olla fysiologinen, sosiaalinen tai kaupallinen. (Bergström & Leppänen 2015, 122.)

Sisäinen ärsyke on jokin henkilön normaali tarve kuten nälkä tai jano, joka ylittää kynnyksen arvon. Myös ulkoinen ärsyke voi herättää tarpeen: kuluttaja ihailee ystävän uutta hyödykettä tai näkee mainoksen, joka inspiroi miettimään ostoa. (Kotler & Keller 2016, 195.) Tällöin hän tulee tietoiseksi, että jokin tuote voi parantaa hänen statustaan tai elintasoaan, jolloin uusi täyttymätön toive herää. Kuluttaja siirtyy prosessissa eteenpäin, kun hän on riittävän motivoitunut ratkaisemaan ongelman. (Bergström & Leppänen 2015, 121–122.) Markkinoijan kannattaa tunnistaa tarpeen laukaisevat olosuhteet keräämällä informaatiota kuluttajista (Kotler & Keller 2016, 195) ja tarjota ratkaisua kuluttajan ongelmaan oikealla hetkellä (Bergström & Leppänen 2015, 122).

2.1.2 Tiedon etsiminen

Kun kuluttaja lähtee ratkaisemaan ongelmaa, hän hankkii tietoa eri vaihtoehtoista (Bergström & Leppänen 2015, 122). Tiedon hakemisessa on kaksi tasoa: kohonnut huomio ja aktiivinen tiedon haku. Ensimmäisessä vaiheessa kuluttaja tulee myötämieliseksi tuotteesta saatavalle tiedolle. Seuraavalla tasolla hän hakee aktiivisesti tietoa eri lähteistä: soittaa ystäville, menee verkkoon, hankkii esitteitä ja vierailee kaupoissa tutustumassa tuotteeseen (kuvio 4). Kuluttajan etsimän tiedon tyyppi ja ylipäänsä vastaanottavuus vaihtelee ajasta ja paikasta toiseen. (Kotler & Keller 2016, 196.)



Kuvio 4. Kuluttajan tietolähteet (mukaillen Kotler & Keller 2016, 196; Bergström & Leppänen 2015, 123. Piirroskuva: Pixabay)

Kerätessään informaatiota hyödykkeistä, kuluttaja oppii kilpailevista brändeistä ja ominaisuuksista. Aluksi kuluttaja tulee tietoiseksi osasta tuotemerkeistä. Harkintaan päätyvät ostokriteerit täyttävät tuotemerkit. Kun kuluttaja hankkii lisää tietoa tuotemerkeistä, vain

muutama säilyy varteenotettavana vaihtoehtona, joista hän tekee lopullisen valinnan. Näin ollen yrityksen tulee saada brändinsä ostajakandidaatin tietoisuuteen, harkintaan ja varteenotettavaksi vaihtoehdoksi. Jos marketissa jogurttimaut asetetaan esille tuotemerkkien mukaan, kuluttajalla on taipumus valita saman tuotemerkin eri makuja. Jos jogurtit on ryhmitelty hyllyyn eri makujen mukaan, kuluttaja valitsee ensin maun ja sitten tuotemerkin. Verkossa etsimiskäyttäytyminen vaihtelee riippuen ainakin osittain tuotetiedon esittämistävasta. Tuotevaihtoehdot voidaan laittaa esille arvioidun houkuttelevuuden perusteella, jolloin kuluttajat voivat valita tuotteet ilman tavanomaista laajempaa etsimistä. (Kotler & Keller 2016, 196–197.) Kuluttaja vertailee eri vaihtoehtojen ominaisuuksia yksinkertaisissakin ostoissa, kuten päivittäistavaroita ostettaessa, koska hän haluaa pienentää ostoon liittyviä riskejä (Bergström & Leppänen 2015, 122–123).

2.1.3 Vaihtoehtojen vertailu

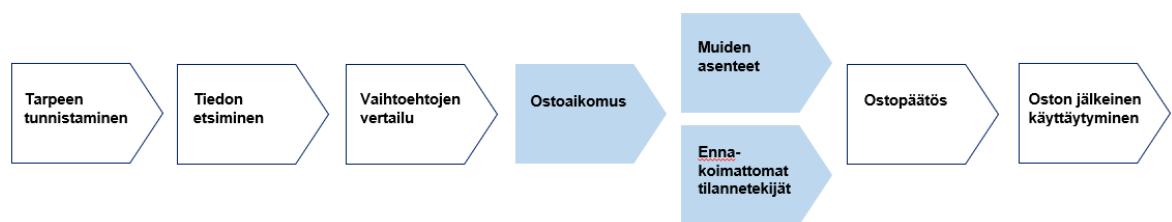
Kuluttajien prosessit tuotemerkkien vertailussa vaihtelevat kuluttajasta toiseen ja eri ostotilanteisiin. Samankaltaisia piirteitä prosesseissa ovat kuluttajan halu tyydyttää tarve, etsiä tuoteratkaisusta tiettyjä etuja ja nähdä ominaisuudet, joilla edut saavutetaan. Vaihtoehtojen vertailuun vaikuttavat myös uskomukset, asenteet ja brändimielikuva. (Kotler & Keller 2016, 197.)

Kuluttaja tekee ostopäätöksen, kun hänellä on riittävä määrä tietoa eri vaihtoehtoista, jotka hän asettaa paremmuusjärjestykseen. Hän voi myös päättää olla ostamatta. Kuluttaja valitsee tuotteet hänelle olennaisten ominaisuuksien tai arvontuoton perusteella. Esimerkkejä valintakriteereistä ovat hinta, status, laatu, takuu, ympäristöystävällisyys tai kotimaisuus. (Bergström & Leppänen 2015, 123.)

2.1.4 Ostopäätös

Ostoaikomusta toimeenpannessaan kuluttaja tekee useita päätöksiä: tuotemerkki, myyjä, määrä, ajoitus ja maksutapa. Kuluttaja voi tehdä ostopäätöksen vertaamalla hyödykkeen hyviä ja huonoja puolia, mutta myös heuristisena valintana, jolloin hän käyttää henkisiä oikeiteitä. Esimerkiksi kuluttaja voi määrittää minimirajan kaikille hyödykkeen ominaisuuksille ja valita ensimmäisen vaihtoedon, joka täyttää vaatimukset. Kuluttaja voi myös valita tuotemerkin tärkeimmän ominaisuuden perusteella. Kolmannessa tavassa kuluttaja vertaa tuotemerkkien tärkeitä ominaisuuksia todennäköisyyksien perusteella ja eliminoi tuotemerkkejä, jotka eivät saavuta minimivaatimuksia. Ostoaikomuksen ja ostopäätöksen välissä vaikuttavat vielä muiden asenteet ja ennakoimattomat tilannetekijät (kuvio 5). Muiden ihmisten, etenkin läheisten, penseys voi vaikuttaa valittuun vaihtoehtoon, samoin jos oma

motivaatio mukautuu muiden toiveisiin. Puolueettomien asiantuntijoiden laatimat kuluttajalähtöiset tuotearviointit vaikuttavat myös kiistattomasti kuluttajien ostopäätöksiin. Ennakoimattomat tilannetekijät voivat muuttaa ostoaikeita esimerkiksi kotitalouden taloudellisen tilanteen heikkeneminen. Mieltymykset ja ostoaikeetkaan eivät ole täysin luotettavia ennustettaessa ostokäyttäytymistä. Kuluttajan päätös muuttaa, lykätä tai välttää ostopäätöstä on usein yhteydessä havaittuun riskiin tai riskeihin. Riski voi olla toiminnallinen, fyysinen, taloudellinen, sosiaalinen, psykologinen ja ajankäyttöön liittyvä. (Kotler & Keller 2016, 198–200.) Myös laatu-, terveydellinen- ja turvallisuusriski mietityttävät kuluttajaa (Bergström & Leppänen 2015, 123–124). Riskin voimakkuus vaihtelee käytettävästä rahamäärästä, epätietoisuudesta ominaisuuksissa ja kuluttajan omasta itsevarmuudesta (Kotler & Keller 2016, 198–200).



Kuvio 5. Vaihtoehtojen vertailun ja ostopäätöksen väliset vaiheet (mukaillen Kotler & Keller 2016, 199)

Käyttäytymispäätösteorian mukaan kuluttajat eivät pysty ennustamaan tulevia mieltymyksiään tarkasti, koska he eivät esimerkiksi tiedä etukäteen miltä tuntuu, kun on syönyt useita kertoja saman makuista jogurttia tai jäätelöä (Kotler & Keller 2016, 203). Löydettyään itselleen sopivan vaihtoehdon kuluttaja ostaa tuotteen. Ostoprosessi voi katketa, jos tuotetta ei ole saatavilla. Ostopäätökseen sisältyy myös valinta verkkokaupan ja kivijalkakaupan välillä, ostopaikan valinta ja ostoehtoista sopiminen. (Bergström & Leppänen 2015, 124.)

2.1.5 Oston jälkeinen käyttäytyminen

Ostoprosessin viimeisessä vaiheessa kuluttaja käyttää tuotetta ja arvioi ostoksena onnistumista. Tyytyväisyys ostoon johtaa todennäköisesti uusintaostoon ja myönteiseen palautteeseen. (Bergström & Leppänen 2015, 124.) Kansainvälinen konsulttiyhtiö McKinsey & Company tutki 125 000 kuluttajan 350 brändin asiakasuskollisuutta 30 kategoriassa. Vain kolmessa kategoriassa kuluttaja eteni ostopolulla uskollisuussilmukkaan eli osti tarpeen ilmetessä uudelleen tietyn brändin vertaamatta sitä muihin brändeihin. Tällaisia kuluttajia oli 13 %. Loput 87 % vertailivat brändejä, ja noin 29 % päätyi lopulta ostamaan vakiintuneen brändin. (Elzinga 2017.) Oston jälkeen kuluttaja tarvitsee päätöstään tukevaa tietoa

tuotteesta. Markkinoijan kannattaakin seurata ja vahvistaa oston jälkeistä tyytyväisyyttä. (Kotler & Keller 2016, 200–201.)

2.2 Kuluttajan omni-channel-ostoprosessi

Kuluttajan omni-channel-ostoprosessi on interaktiivinen ja moninainen. Kaupankäyntikanavia ovat kivijalkamyymälä, verkkosivut, suoramarkkinointi, mobiilikanavat, esimerkiksi älypuhelimet ja sovellukset, ja sosiaalinen media. Kuluttajan kontaktipisteisiin kuuluvat massakommunikaatiokanavat kuten TV, radio ja printtimedia. Erilaiset integroidut kanavat muodostavat omni-channel-toimintatavassa saumattoman asiakaskokemuksen. (Verhoef, Kannan & Inman 2015.) Kuluttaja valitsee omni-channel-ostoprosessin, kun teknologian käyttäminen ja kanavien yhdistäminen tarjoaa merkittävää etua, kuten esimerkiksi nopeampaa ja helpompaa ostamista (Susanto, Sucahyo, Ruldeviyani & Gandhi 2018). Kun kaupankäynti siirtyy digikanaviin kasvaa sosiaalisen median ja brändien merkitys. Tuttu ja kuluttajalle arjessa läheinen brändi korreloi seuraavan ostopäätöksen kanssa. Kuluttajan ostoprosessi on usein epälineaarinen, dynaaminen ja verkottunut. (Palokangas 2019.)

Vuonna 2019 tehdyn vähittäiskaupan ostotutkimuksen (n = 7 000) mukaan 67 % suomalaisista kuluttajista aloittaa ostopolun verkossa. Ostajat jakautuivat neljään tyyppiin ostopolkunsa perusteella. Eniten eli 46 % kuluttajista haki tietoa digikanavista, mutta osti tuotteen kivijalkakaupasta. Seuraavaksi eniten, 31 % kuluttajista, aloitti ja päätti ostopolkunsa fyysisessä kaupassa. Verkossa tietoa hakevien ja ostavien kuluttajien osuus oli 21 %. Heidän määränsä on lähes kaksinkertaistunut kahdessa vuodessa. Ainoastaan 1 % kuluttajista haki tietoa tuotteesta kivijalkakaupasta, mutta osti tuotteen netistä. (Kuulas Helsinki 2019.)

Verrattaessa kuluttajien (n = 30 000) päivittäistavaroiden ostopäätösten ajoituksia Pohjoismaissa vuonna 2019 havaittiin, että enintään 6 kuukautta aikaisemmin ostopäätöksen teki 6 % kuluttajista ja 1–5 kuukautta aikaisemmin 28 % kuluttajista. 58 % kuluttajista teki ostopäätöksen lähempänä: 14 % teki ostopäätöksen 2–3 viikkoa aikaisemmin, 20 % edellisellä viikolla, 16 % samana päivänä ja 8 % juuri ennen ostosta. (Kuulas Helsinki 2019.)

3 Ulkomainonta liiketoimintana

Kohdeyrityksen toimiala on kodin ulkopuolinen mainonta, jolloin yritys myy mediaan näkyvyyttä mainostaville yrityksille, jotka tavoittelevat kuluttaja-asiakkaiden huomiota. Ulkomainonnalla ja digitaalisella ulkomainonnalla tarkoitetaan kaikkea kodin ulkopuolista mediaa, kuten esimerkiksi kauppakeskuksissa, rakennusten ulkoseinissä ja kadun varsilla sijaitsevia mainospintoja. Mainospinnat vaihtelevat painetuista julisteista LED-pinoihin. (Mediateko 2020.) Ulkomainonnan kysyntä on kasvanut tasaisesti viimeiset viisi vuotta Suomessa, ja sen arvo oli 66,3 miljoonaa euroa vuonna 2018. TNS Kantarin tutkimuksen mukaan kasvua edelliseen vuoteen tuli 15 % vuonna 2018. (Mediateko 2020.) Ulkomainontaa kutsutaan viimeiseksi massamediaksi, koska se tavoittaa kaikki vaikutuspiirinsä ihmiset sukupuoleen ja ikään katsomatta (Riipinen 2019).

Ulkomainontaa sijoitetaan kuluttajan ostopolun varrelle, sillä päivittäistavaroiden valmistajat ja markkinoijat ovat tärkeitä mainostajia. Päivittäistavarakaupan kokonaismyynti Suomessa oli 18,2 miljardia euroa vuonna 2018. Suurin osa päivittäistavaramyynnistä tapahtuu kivijalkamyymälöistä. Verkkokaupan osuus myynnistä vuonna 2018 oli vain 70 miljoonaa euroa (0,4 %), mutta trendi on kasvava. (Nielsen 2019.) Kevään 2020 koronavirusepidemia on jopa lähes nelinkertaistanut päivittäistavaroiden verkkokaupan tilausmäärät (Elonen 2020).

3.1 Ulkomainonta massamediavälineenä

Ulkomainonta on massamediavälineenä hyvin monipuolinen. Se on kohdennettavissa kohderyhmittäin sekä alueen, paikan, ajan ja säätilan mukaan. (Mediateko 2020.) Digitaalinen ulkomainonta on ketterää, jolloin mainosviesti voidaan kohdistaa oikeassa paikassa oikeaan aikaan, kuten jäätelömainos kesähelteellä (Ylä-Anttila 2019). Tavoittavuus on ulkomainonnan suurin vahvuus: kun ulkomainonta tavoittaa viikoittain jopa 97 % suomalaisista, on Internetin vastaava luku 96 %. Digitaalisen ulkomainonnan tuotantokustannukset ovat edulliset, päivittäminen on nopeaa ja personointi joustavaa esimerkiksi ajallisesti. Nielsenin tutkimuksen mukaan ulkomainonta ohjaa paremmin verkkoon kuin televisio, radio tai sanomalehti. (Mediateko 2020.) Outsmartin selvityksessä se lisäsi myös brändin mobiililiikenteen kasvua keskimäärin 17 %. Nuorten keskuudessa mobiiliaktiivisuus lisääntyi yli 140 %. (JCDecaux 2020a.) Ohjelmallinen ostaminen eli mediatilan ostaminen automaattisesti järjestelmien kautta on mahdollista verkkomedioiden ja digikanavien ohella myös ulkomainonnassa. Sen etuna on datan hyödyntäminen kohdentamisessa ja kampanjan reaaliaikainen optimointi. (Kukkonen 2018.)

Ulkomainontaa mitataan Outdoor Impact -yleisönmittausjärjestelmällä. Menetelmässä yhdistyvät kuluttajien kulkemat reitit ja liikennevirrat kaduilla ja teillä. (Clear Channel 2020a.) Mediavaluutta VAC ilmaisee todennetut katsekontaktit eri ulkomedioissa ja mediayhdistelmissä havainnointitutkimuksen mallinnusmenetelmän mukaan (Outdoor Impact 2020). Kun Lumen Researchin tutkimuksessa verrattiin eri medioiden välisiä eroja arvioitaessa yleisön huomiota, eri mittareilla nähtyjen mainosten (VAC) osuus oli keskimäärin tietokoneella 13 %, mobiilissa 25 %, klassisessa ulkomainonnassa 66 % ja digitaalisessa ulkomainonnassa 39 % (Follet 2020).

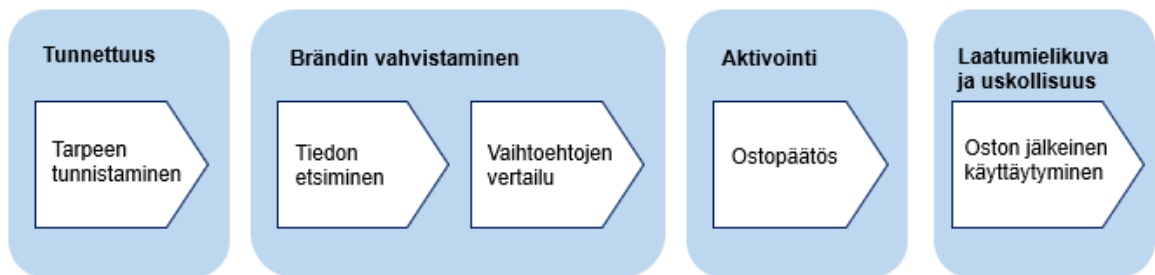
Digitaaliset mainosnäytöt mahdollistavat interaktiivisen ja dynaamisen, esimerkiksi ajankohdan mukaan elävän, mainossisällön. Tällöin esimerkiksi aamiaistarvikkeita voi mainostaa aamulla ja illallisraaka-aineita illalla. (Clear Channel 2020b, 14.) Ulkomainonnassa mainostettavalle tuotteelle löydetään paras yleisö datan avulla. Kun tunnistetaan, missä tuotteesta kiinnostuneet kuluttajat liikkuvat, voidaan kampanja kohdistaa tehokkaasti. Kohderyhmä on mahdollista löytää esimerkiksi sijaintihistorian mukaan. Kuluttaja on haluttua kohderyhmää, jos hän vierailee usein tuoteryhmään liittyvissä toimipisteissä. Mainosten parhaat sijaintipaikat arvioidaan pisteytyksellä yhdistämällä sijaintiin kohderyhmän osuus, esiintymisuseus ja relevanssi, ja lopuksi ajankohta. Tavoitteena on, että kampanjan voisi kohdistaa suoraan yleisölle, eikä kohdentamista sijainteihin tarvittaisi. (JCDecaux 2019c.)

3.2 Ulkomainonnan vaikutus kuluttajan ostopäätösprosessiin

Kuluttajan kaupassakäynti kestää Suomessa keskimäärin 30 minuuttia. Myymälässä hänen huomiostaan kilpaillaan muun muassa tuotteiden esillepanolla, pakkauksilla, tuotesittelyillä, mainonnalla ja hinnalla. Hintaviestejä kuluttaja näkee jopa 500 eli kaksi jokaista sekuntia kohden (Clear Channel 2020b, 4, 13).

Myymälämediaverkoston mainosnäytöt on sijoitettu kauppojen sisääntuloväylille, ikkunoihin ja sisälle myymälöihin muun muassa keskeisille kulkuväylille (Clear Channel 2020b, 12). Niiden avulla voidaan vaikuttaa kuluttajan ostopäätökseen monin tavoin eri ostopolun vaiheissa (kuvio 6). Kun kuluttaja on matkalla vähittäismyymälään ostoksille, voidaan mainonnalla vaikuttaa hänen tietoisuuteensa tuotteesta ja herättää ostokiinnostus. Harkintavaiheessa mainonta auttaa erottautumaan muista tuotteista ja ostopäätös vaiheessa poimimaan ostoskärriin tietyn tuotteen, sillä 67 % kuluttajista päättää tietyn brändin ostosta vasta myymälässä. (Clear Channel 2020b, 9–10.) Ulkomainonnalla päästään muita medioita lähemmäksi ostopäätös vaiheessa (JCDecaux 2020a). Ostopaikan lähelle sijoitettu POS-mainonnan tarkoitus on vaikuttaa brändin valintaan ja ostoskorin koostumukseen

(Riipinen 2019). Suurin osa ostopäätöksistä tehdään vasta myymälässä, ja myymälämainokset vaikuttavatkin 63 % suomalaisen kuluttajan päivittäistavaroiden ostopäätöksiin Kantar Mindin mukaan (Clear Channel 2020b, 8–9). Esimerkiksi meijerituotteiden valmistaja sai nostettua kampanjaviikon myyntiä 42 % tavalliseen viikkoon verrattuna myymälämainonnan avulla (Clear Channel 2020b, 13). Kuluttajat huomasivat myymälöiden digitaalisista elementeistä toiseksi parhaiten digitaaliset näytöt (71 %) heti videoseinäkkeiden (73 %) jälkeen. Kaupan interaktiivinen teknologia voi tukea kuluttajaa ostopäätösprosessissa etenkin tuoteinformaation löytämisessä ja tuotteiden sijainnin helpommassa paikantamisessa. Iso-Britanniassa ja USA:ssa vähittäiskaupat ovat saavuttaneet 3 %:n kasvun liikevaihtoon teknologian avulla. (Azhari & Bennett 2015.)



Kuvio 6. Ulkomainonnan vaikutus kuluttajan ostopäätösprosessin eri vaiheissa (mukaillen JCDecaux 2020b)

Kuluttajat toimivat kaupassa joko suunnitelmallisesti tai impulsiivisesti. Suunnitelmallinen ostaja on päättänyt ostaa jonkin tuotekategorian tuotteen. Ärsykkeen avulla tietty tuote saadaan erottumaan massasta brändimielleyhtymän avulla. Impulsiivisessa ostajassa herätetään ärsykkeen avulla tarve tiettyä tuotetta kohtaan ja ohjataan hänet tuotteen luo. Ostajiin vetoava mainos on tehokas ärsyke, joka voi johtaa ostoon. (Clear Channel 2020b, 6.)

4 Tutkimusmenetelmien ja työkalujen kuvaus sekä teoreettinen tausta

Tässä luvussa kuvataan käytetyt tutkimusmenetelmät ja työkalut sekä perustellaan valinnat teoriaan perustuen.

4.1 Puolistrukturoitu haastattelu

Haastattelu valittiin aineistonkeruumenetelmäksi, koska haluttiin saada taustatietoa yrityksen liiketoiminnasta ja liiketoimintaympäristöstä ratkaistavan ongelman tunnistamiseksi.

Haastattelu sopii ilmiöiden tutkimiseen, kun tutkimuskohdetta ei vielä tunneta niin hyvin, että yksityiskohtaisia kysymyksiä voitaisiin esittää (Kananen & Makkonen (toim.) 2015, 143). Haastattelumenetelmäksi valittiin puolistrukturoitu haastattelu, koska se soveltuu käytettäväksi juuri sellaisessa tilanteessa, kun tutkimuksen kohdetta ei tunneta hyvin etukäteen. Tällöin teemat mietitään etukäteen, mutta kysymysten järjestys, sanamuodot ja painotus voi vaihdella. (Ojasalo ym. 2014, 41.) Haastattelun toteutus on kuvattu kohdassa 5.1.

4.2 Dokumenttianalyysi

Dokumenttianalyysi valittiin aineistonkeruumenetelmäksi, koska haluttiin perehtyä kehittämisen kohteeseen keräämällä tietoa ilmiöstä ja selvittämällä, miten datan avulla voi lisätä kuluttajaymmärrystä.

Dokumenttianalyysi tuottaa erilaisia näkökulmia kehittämisen kohteesta. Sitä käytetään yleensä yhdessä muiden tiedonkeruumenetelmien kanssa. Kehittämisen tavoitteiden täsmällisen määrittämisen takia on tärkeä perehtyä muun muassa kohdeyrityksen taustatietoihin, toimialaan ja toimintaympäristöön. (Ojasalo ym. 2014, 28–29, 43.) Dokumenttianalyysin toteutus on kuvattu kohdassa 5.1.

4.3 Elintarvikenimikkeiden määrittäminen COICOP-HBS-luokituksella

Tutkittaviksi tapauksiksi valitut neljä elintarviketuoteryhmää, meijerituotteet, valmisruoat, kasviproteiinit ja sesonkituotteet, haluttiin purkaa järjestelmällisesti elintarvikenimikkeiksi, jotta yksittäisten tuotteiden hakemista voitaisiin tutkia. Ensisijaiseksi menetelmäksi valittiin Tilastokeskuksen kulutustutkimuksen hyödykeluokituksessa käyttämä kansainvälinen COICOP-HBS-kulutusluokitus (yksilöllisen kulutuksen käyttötarkoituksen mukainen luokitus) (Tilastokeskus 2019a), josta on laadittu kansalliset erityispiirteet huomioiva versio. Se

kattaa yli 280 elintarvikenimikettä. Elintarvikkeiden luokittelu perustuu pääraaka-aineseen, ja edelleen niiden alaryhmiin. Useita raaka-aineita sisältävät valmisruoat luokitellaan vallitsevan ainesosan mukaan. (Aalto 2018, 9.)

Meijerituotteilla tarkoitetaan maitopohjaisia tuotteita. Maitotuotteet luokitellaan COICOP-luokituksella neljään pääluokkaan: maito, jogurtti, juustot ja rahkat sekä muut maitotalousvalmisteet (Tilastokeskus 2019b).

Valmisruoat luokitellaan COICOP-luokituksella kolmeen pääluokkaan: valmisateriat, osan aterialaajaa muodostavat valmisruoat ja puolivalmisteet. Valmisaterioita ovat valmisruoat, jotka muodostavat kokonaisen aterian ja syödään sellaisenaan lämmittämisen jälkeen. Ne luokitellaan neljään alaryhmään, joista tässä tutkimuksessa tarkastellaan laatikot, keitot, salaattit ja pasta-ateriat -ryhmää. (Aalto 2018, 10–11.)

Kasviproteiineista valmistetut tuotteet luokitellaan COICOP-luokituksella kasvipihvien ja -salaattien ryhmään. Aterian osan muodostavia valmisruokia ovat esimerkiksi soijanakit, seitanmakkarat, Härkis-pyörykät ja Nyhtökaura-pihvit. Kasviproteiiniiniryhmän luokka on 0117608, Tofu. (Aalto 2018, 11.) COICOP-luokitukselta ei löydy kuitenkaan kattavaa alaluokitusta kasviproteiiniiniryhmän tuotteille, joten niiden luokitus perustuu www.vegaanituotteet.net -sivuston tuotevalikoimaan.

Sesonkituotteissa vuoden tärkeimpien sesonkien ajankohdat ja tyypilliset tuotteet kartoitettiin vuosien 2018 ja 2019 perusteella Tilastokeskuksen tilastokirjastojen (Tilastokeskus 2019d, 2019e, 2019f & 2018a) ja Ubersuggest-palvelun avulla 28.11.2019 ja koottiin taulukoksi (liite 2). Sesongeista valittiin jouluku syvällisempää tarkastelua varten, koska se osui tutkimuksen tekemisen aikajänteelle, ja näin tutkittavista tuotteista pystyttäisiin keräämään Google Trends -dataa tuntien tarkkuudella.

Elintarvikenimikkeiden määrittäminen on kuvattu kohdassa 5.3.1.

4.4 Hakutrendien analysointi

Hakukoneanalytiikan hakutrendien analysointi valittiin menetelmäksi, koska tutkimuksessa haluttiin selvittää, milloin kuluttajat hakevat tietoa elintarvikeryhmien nimikkeistä, mitä hakusanoja he käyttävät haussa ja kuinka paljon eri hakutermejä haetaan. Päivittäiselintarvikkeiden tuotenimikkeiden hakuajankohdat sijoittuvat ostopäätösprosessissa tiedon etsiminen -vaiheeseen.

Maailman suosituin hakukone on Google 92 %:n markkinaosuudella (Statcounter 2020a) ja toiseksi suosituin on YouTube, sillä kolmasosa nettiajasta käytetään videoiden katseluun (HubSpot 2020). Tutkimuksessa keskitytään Googlen verkkohakuihin, koska se on ylivoimaisesti käytetyin hakukone Suomessa 96,4 %:n markkinaosuudella, kun kaikki alustat lasketaan mukaan (Statcounter 2020b). Google-hakuja voi lisäksi analysoida avoimella Google Trends -työkalulla. Lisäksi hakutrendien analysointi katsottiin tarpeelliseksi, koska esimerkiksi Boonen ym. (2015) mukaan hakutermien valinta voi olla haasteellista sisällytettäessä Google Trends -dataa ennustemalliin. Toisaalta hakukoneoptimoinnissa hakusanojen valinnalla on suurin vaikutus toimenpiteiden tehokkuuteen (Sponder & Khan 2018, 28).

Hakutrendejä analysoitaessa selvitetään aluksi, millä hakusanoilla ja -lauseilla ihmiset etsivät tietoa (Suojanen 13.7.2016). Avainsanoja tutkittaessa selvitetään, mitä ihmiset hakevat (informaatio, tuotteet, palvelut), millaisilla hakulausekkeilla ja kuinka paljon (Jukarainen 20.11.2019). Google Trendsillä ei näe hakujen määrää, mutta hakuvolyymit voi selvittää Ubersuggest-työkalulla.

Hakutrendien analysoinnin toteutus on kuvattu kohdassa 5.3.1.

4.5 Kahden muuttujan välisen riippuvuuden tutkiminen

Tutkimuksessa ollaan kiinnostuneita punaviinin myynnin ja Google Trends -hakuindeksien välisestä riippuvuudesta, jotta niille voidaan valita sopivat jatkoanalyysimenetelmät ennustamiseen. Aluksi selvitetään, mitkä menetelmät on mahdollista valita riippuvuuden tutkimiseen nelivaiheisen mallin avulla.

Kahden muuttujan välistä riippuvuutta voi tutkia neljän toisiaan seuraavan vaiheen avulla. Ensin tehdään järkevyytarkastelu eli tarkastellaan, ovatko muuttujat luonteeltaan sellaisia, että niillä on tekemistä keskenään vai onko niiden taustalla vaikuttamassa jokin kolmas muuttuja. Jos muuttujien yhteys vaikuttaa järkeenkäyvältä, siirrytään seuraavaan vaiheeseen valitsemaan analysointimenetelmä. Se valitaan muuttujien mitta-asteikon tason ja jakaumien muodon perusteella. Kolmannessa vaiheessa tehdään hajontakuviot. Viimeisessä vaiheessa valitaan matemaattiset mallit. Esimerkiksi korrelaatiolla voidaan tarkastella riippuvuuden voimakkuutta ja regressioanalyysillä tehdä ennustemalli. (Saaranen 2018, 55.)

Riippuvuusanalyysin valinta tehdään muuttujien mitta-asteikon perusteella. Muuttuja merkitsee tutkittavaa ominaisuutta. Jos se on tyypiltään kvantitatiivinen eli määrällinen vastaus ilmaistaan numeerisena arvona. (Saaranen, 2018, 3.) Määrälliset muuttujien arvot voivat olla välimatka- tai suhdeasteikollisia (Saaranen 2018, 6). Välimatka-asteikolla mitataan numeerisia muuttujia, jotka voivat saada mitä tahansa reaalilukuarvoja. Havainto ilmoitetaan sen suuruutta kuvaavalla lukuarvolla, jonka etäisyys voidaan kertoa. (Nummenmaa 2011, 42.) Suhdeasteikolliset muuttujat voi ilmaista nimensä mukaisesti suhteina: A on kaksi kertaa suurempi kuin B esimerkiksi litrat ja prosenttiosuudet (Saaranen 2018, 6).

Jos sekä selittävä että selitettävä muuttuja ovat välimatka- tai suhdeasteikollisia voidaan valita analysointitavaksi Pearsonin korrelaatio, hajontakuviot, selityssaste sekä regressioanalyysi, mikäli muuttujien välillä ilmenee korrelaatiota (Saaranen 2018, 54). Muuttujien jatkuvuus mahdollistaa Pearsonin korrelaatiokertoimen valinnan kuvaamaan riippuvuuden voimakkuutta (Kananen & Makkonen (toim.) 2015, 345). Muuttujien jatkuvuus tarkoittaa, että mittaustuloksena saatu arvo voi olla mikä tahansa reaaliluku muuttujien arvojen vaihteluväliltä (Nummenmaa 2011, 40). Ominaisuus on ilmaistavissa desimaaleja käyttäen (Saaranen 2018, 6).

Hajontakuvioista tulee testata riippuvuuden lineaarisuus, sillä Pearsonin korrelaatiokerroin olettaa, että riippuvuus on lineaarinen (Kananen & Makkonen (toim.) 2015, 345). Regressioanalyysin muuttujien lähtöoletukset ovat samat kuin korrelaatioanalyysillä (Kananen & Makkonen (toim.) 2015, 316).

Kahden muuttujan välisen riippuvuuden tutkiminen on kuvattu kohdassa 5.3.3.

4.6 Ristikorrelaatio- ja spatiotemporaalinen analyysi

Ostoprosessin ostopäätösvaihetta voidaan tutkia Google-hakujen ja ostamisen välisellä viiveellä. Ristikorrelaatioanalyysi valittiin tutkimusmenetelmäksi, koska viivelukujen avulla voidaan selvittää muuttujien eli punaviinin Google Trends-hakuindeksien ja myynnin välinen aikajänne Suomessa ja maakunnissa. Tutkimuksen muuttujat sopivat ristikorrelaatioanalyysiin, koska molemmat ovat jatkuvia (kohta 4.5) ja niistä on saatavana katkeamaton kuukausidata koko tarkasteltavalta ajanjaksolta. Spatiotemporaalisella analyysillä visualisoidaan viive-erot eri maakunnissa ja ajanjaksoina kartalle.

Ristikorrelaatioanalyysi on yksinkertainen menetelmä tutkia kahden aikasarjan välistä riippuvuutta samanaikaisesti ja erilaisilla aikaviiveillä (Holmes, Scheuerell & Ward 2020). Muuttujien arvot voivat liikkua samanaikaisesti tietyllä ajanhetkellä tai toisen muuttujan

liike voi edeltää tai seurata toisen liikkeitä. Ristikorrelaatio lisää ymmärrystä muuttujien välisestä riippuvuudesta ja liikkeistä, ja sitä käytetään testaamaan aikasarjamuuttujien keskinäisiä vaikutuksia. (IBM 2020a.) Kinskin (2016) mukaan Google Trends -datan ja myyntidatan käyttäminen ristikorrelaatioanalyysissä soveltuu ennustamaan viivettä Google-hakujen ja uusien autojen myynnin välillä.

Spatiotemporaalisella analyysillä tarkoitetaan uuden tiedon tuottamista käsittelemällä tiettyyn paikkaan tai alueeseen viittaavaa tietoa ja tähän sisältyvää aikatietoa (Sanastokeskus TSK 2020).

Analyysien toteutus on kuvattu kohdassa 5.3.2.

4.7 Pearsonin korrelaatioanalyysi

Aluksi tutkittiin kahden muuttujan välistä riippuvuutta (kohdat 4.5 ja 5.3.3.), jonka perusteella valittiin Pearsonin korrelaatioanalyysi menetelmäksi tutkia punaviinin myynnin ja Google Trends -hakuindeksien välistä korrelaatiota Suomessa. Käytettäessä Internetiä ostopäätöksen tutkimiseen lähtökohtana on, että termien hakeminen enteilee vastaavien tuotteiden tai tuotekategorioiden myyntiä. Tällöin hakutermit merkitsevät aikomusta ostaa tiettyjä tuotteita, jotka korreloivat voimakkaasti hakutermin kanssa. (Boone ym. 2015.) Korrelaatio soveltuu käytettäväksi, kun halutaan mitata kahden muuttujan välistä riippuvuutta ja sen voimakkuutta ja suuntaa, joka voi olla positiivinen tai negatiivinen (Kananen & Makkonen (toim.) 2015, 314–315).

Korrelaatiokerroin

Korrelaatiokerroin kertoo suoraviivaisen riippuvuuden voimakkuudesta. Sen arvo vaihtelee -1 ja +1 välillä. Jos arvo on lähellä nollaa, ei suoraviivaista riippuvuutta ole havaittavissa, arvo lähellä +1:tä kertoo positiivisesta riippuvuudesta ja arvo lähellä -1:tä vastaavasti negatiivisesta riippuvuudesta. (Taanila 17.4.2019.)

Korrelaatiokertoimia tulkitaan Saarasen (2018, 62) mukaan seuraavasti:

- Voimakas positiivinen lineaarinen riippuvuus (0,7–1)
- Kohtalainen positiivinen lineaarinen riippuvuus (0,3–0,7)
- Ei lineaarista riippuvuutta (-0,3–0,3)
- Kohtalainen negatiivinen lineaarinen riippuvuus (-0,3–(-0,7)).
- Voimakas negatiivinen lineaarinen riippuvuus (-0,7–(-1)).

Esimerkiksi Kananen & Makkonen (toim.) (2015) käyttävät itseisarvojen raja-arvoina 0,7 sijaan 0,8 ja 0,3 sijaan 0,4, mutta tässä tutkimuksessa korrelaatioita tulkitaan Saarasen arvojen mukaan, joita myös Taanila käyttää. Lisäksi Taanilan mukaan, jos otoskoko on

suuri, itseisarvoiltaan pienemmätkin kertoimet ovat tulkittavissa merkitseviksi. (Taanila 2019a, 39.)

Selitysaste

Korrelaatiokertoimen selitysasteesta näkee selittävän muuttujan (x) kyvyn kuvata selitettävän muuttujan (y) vaihtelua. Selitysaste on korrelaatiokerroin korotettuna toiseen potenssiin. Tulos esitetään usein prosenttilukuna. (Saaranen 2017, 56.)

Pearsonin korrelaatioanalyysin toteuttaminen on kuvattu kohdassa 5.3.3.

4.8 Regressioanalyysi

Kahden muuttujan välisen riippuvuuden tutkimisen (kohdat 4.5 ja 5.3.3.) tuloksena valittiin korrelaatioanalyysin jatkoanalyysiksi regressioanalyysi. Regressioanalyysi sopi menetelmäksi, koska haluttiin ennustaa punaviinin myyntimääriä selittävän muuttujan eli Google Trends -indeksien avulla ja hyödyntää aikasarjojen historiatietoja toteutuneista myynneistä.

Regressioanalyysi on käyttökelpoinen menetelmä esimerkiksi myynnin ennustamiseen ja selittävien muuttujien painoarvon arvioimiseen, kun halutaan tutkia eri osa-alueiden vaikutusta kokonaisuuteen (Taanila 2010, 10) ja aikasarjojen tutkimiseen (Sharda 2018, 123). Myynnin ennustamista voi kehittää menneiden myyntien pohjalta, sillä aikasarja-analyysillä voidaan tutkia toteutuneiden aikasarjojen perusteella trendiä, sykliä, kausiluoteisuutta sekä arvaamattomuutta, epävarmuutta ja säännöttömyyttä sekä arvioida tulevaa kehitystä (Kotler & Keller 2016, 115).

Aikasarjoihin perustuvassa ennustamisessa lähtökohta on, että toteutuneet havainnot auttavat tulevien havaintojen ennustamisessa. Ennustusmenetelmä valitaan aikasarjan systemaattisen vaihtelun perusteella. (Taanila 22.12.2019.) Koska punaviinin pitkän aikavälin Google Trends -hakutuloksissa näkyy trendi ja selkeä kausivaihtelu joulun seudulla, olisi ennustetta mahdollisuus parantaa korjaamalla trendiä kausivaihtelulla, jolloin menetelmä sopisi Holt-Winterin kolminkertainen eksponentiaalinen tasoitus (Taanila 22.12.2019). Toisaalta trendi on helppo sovittaa myös lineaariseen autoregressiiviseen malliin, mutta aikasarjojen käänkökohdat hankaloittavat ennusteen tekemistä (Choi & Varian 2011). Sinällään lineaarinen regressiomalli sopii ennusteiden laskemiseen, jos trendikin on lineaarinen (Taanila 2015, 19).

Regressiomenetelmät soveltuvat muuttujien välisten yhteyksien mallintamiseen ja mittaus-tulosten ennustamiseen (Nummenmaa 2011, 309). Kun halutaan saada selville muuttujien välisen riippuvuuden tarkempi matemaattinen muoto, tulee etsiä hajontakuvioiden pistepar-vea kuvaava suora, jonka jäännösarvot ovat pieniä. Regression korkea korrelaatioarvo kertoo muuttujien voimakkaasta riippuvuudesta. (Kananen & Makkonen (toim.) 2015, 316–317.) Kahden muuttujan riippuvuutta voidaan kuvata rakentamalla matemaattinen malli, jossa toisen muuttujan arvoista voidaan laskea toisen muuttujan arvot. Mallina käytetään suoraa, jos muuttujien välillä on lineaarinen riippuvuus. (Taanila 2015, 11.) Mallin-tamisessa selitetään selitettävän muuttujan y arvojen vaihtelua selittävällä muuttujalla x , jolloin saadaan matemaattinen yhtälö kuvaamaan muuttujien välistä yhteyttä. Kun x tun-netaan, voidaan ennustaa y tai toisinpäin. (Nummenmaa 2011, 309.) Esimerkiksi, jos pu-naviinin hakuintensiteetin ja myynnin välillä on tunnettu riippuvuus, voidaan arvioida kuinka paljon 5 %:n kasvu hakuintensiteetissä lisää myyntiä (Kananen & Makkonen (toim.) 2015, 316–317). Mallin laskemisessa mukana olevat havainnot rajaavat käyttöalu-een, koska muuttujien välinen riippuvuus havaintoalueen ulkopuolella ei ole tiedossa (Taanila 2010, 10).

Beta-kerroin on regressiosuoran kulmakerroin normitetuille arvoille. Tämä tarkoittaa, että arvot on standardoitu samalle asteikolle, jotta muuttujien skaala saadaan yhtenäistettyä keskinäisen selityskyvyn vertailun parantamiseksi. (Saaranen 2018, 71.) Tämä on tärkeää sen vuoksi, että muutoin selittävien muuttujien painoarvoa olisi vaikea nähdä, jos muuttu-jien arvojen suuruusluokka on erilainen. Jos selittävä muuttuja kasvaa yhdellä keskihajon-nalla, niin selitettävä muuttuja kasvaa beta-kertoimen arvon verran, mikäli muut selittävät muuttujat pysyvät ennallaan. (Taanila 2010, 10). Jos muuttujia on vain kaksi, ovat beta-kerroin ja Pearsonin korrelaatiokerroin samat (Saaranen 2018, 71). Mitä suurempi selitys-kerroin, sitä paremmin malli selittää selittävän muuttujan vaihtelua. Mutta selityskerroin ei yksin kerro mallin toimivuudesta käytännössä, vaan siihen tarvitaan hyvää ilmiön ymmär-tämistä, jonka perusteella selittävät muuttujat on valittu. (Taanila 2010, 11.)

Google Trends -hakudatan avulla voi parantaa vähittäiskaupan tuotteiden myynnin ennus-temallin tarkkuutta. Kysynnän ennustamisen perustuessa ainoastaan myynnin historiada-taan, on riskinä voittojen sulaminen, jos käsiin jäävä tavara joudutaan myymään alehin-nalla tai tavaran loppuessa kesken jää tuloja saavuttamatta. Yrityksen sisäistä ja ulkoista dataa tuotteista, prosesseista ja asiakkaista voidaan käyttää myyntidatan tukena paranta-maan ennustuksia ja toiminnan tehokkuutta. Datan hyödyntämisen haasteita ovat kuiten-kin datan tuottamisen reaaliaikaisuus, suuri nopeus ja volyymi sekä lukemattomat variaa-tiot. (Boone ym. 2015.)

Google-hakuindeksit korreloivat usein erilaisten talouden indikaattoreiden, kuten esimerkiksi myynnin kanssa, ja niitä voi käyttää lyhyen aikavälin talousennusteisiin (Choi & Varian 2011.) Google-hauista voi ennustaa jopa osakemarkkinoiden kehitystä, sillä odottamaton kasvu Googlen [bear market] (karhumarkkina) -hakutermissä kertoo piensijoittajien pessimismistä ja ennustaa negatiivisia tuottoja seuraavalle viikolle. Google-haut kuvaavat perinteisiä kyselyitä oikea-aikaisemmin piensijoittajien tuntemuksia tulevista tuotoista lyhyellä aikavälillä. (Klemola 2019, 77–78, 104–105.) Vaikka hakudatan käytöllä ei voitaisiakaan kasvattaa mullistavasti ennustettavuutta, voidaan taloudellisesti merkittäviä parannuksia tehdä ennustuksen suorituskykyyn. Toisaalta hakutulokset ovat erityisen käyttökelpoisia ostoksissa, joiden ostamista kuluttajat alkavat suunnitella huomattavasti ennen ostopäätöksen tekemistä. (Choi & Varian 2011.)

Regressiomallin onnistumisen arviointi

Regressiomallin onnistumista voidaan arvioida mallin sopivuudella ja selitysasteella, selittäjän sopivuudella ja jäännöstermeillä (Nummenmaa 2011, s. 320).

Regressioanalyysi ennustaa selitettävän muuttujan arvoja selitettävillä muuttujilla, jolloin mallilla voidaan selittää osuus vaihtelusta. F-suhde kertoo mallin sopivuudesta selitetyn riippuvan muuttujan vaihtelun ja jäännöstermien suhteessa. Mallin katsotaan sopivan aineistoon, jos tulos on tilastollisesti merkitsevä ($p < 0,05$). (Nummenmaa 2011, 320.)

Mallin selitysaste kuvaa, miten suuri osa selitettävän muuttujan vaihtelusta on mahdollista kuvata mallilla. Selitysaste vaihtelee 0 ja 1 välillä, koska se perustuu korrelaatiokertoimen neliöön lineaarisessa mallissa. Selitysasteella 1 on kaikki selitettävän muuttujan vaihtelu kuvattavissa mallilla. (Nummenmaa 2011, 320–321.)

Jos selittävä muuttuja sopii hyvin malliin, viittaa se muuttujan kyvykkyyteen selittää selitettävää muuttujaa (Taanila 2010, 10). Regressiokertoimesta näkee selittävän muuttujan merkityksen y-muuttujan vaihtelulle. Positiivisella x-muuttujan kertoimella arvojen kasvaessa kasvavat myös y-muuttujan arvot. Kertoimen ollessa negatiivinen x-arvojen kasvu pienentää y-muuttujan arvoja. Y-muuttujan vaihtelua voidaan ennustaa sitä enemmän, mitä suurempi on itseisarvoltaan x-muuttujan regressiokerroin. Jos muuttujan regressiokerroin on arvoltaan lähellä nollaa, muuttujalla ei ole juuri merkitystä mallille. (Nummenmaa 2011, 321–322.)

Jäännöstermi kuvaa regressiomallissa esiintyvää virhettä (Saaranen 2017, 58). Se on mallin ennustaman ja toteutuneen havaintoarvon erotus (Nummenmaa 2011, 329) y-

muuttujan arvoille (Nummenmaa 2011, 310). Jäännös kuvaa siis y :n arvon osaa, jota mallilla ei voi ennustaa. Jäännöstä kutsutaan myös jäännösvirheeksi, koska se kuvaa juuri mallin virhettä. (Saaranen 2017, 58.)

Regressioanalyysissä malli toimii y -muuttujan tarkassa ennustamisessa, kun jäännöstermien itseisarvot ovat mahdollisimman pieniä. Mallin korjatusta selitysteesta saadaan arvio mallin selitysvoimalle. Hyvin toimivassa mallissa jäännöstermit jakautuvat satunnaisesti tai normaalisti. Tällöin malli toimii samalla tavalla kaikilla y -muuttujan arvoilla, kun yhteys havaitun arvon ja jäännöstermin väliltä puuttuu. (Nummenmaa 2011, 324.) Jäännöstermi 0 tarkoittaa, että havaittu ja ennustettu arvo ovat samoja, jolloin malli on toiminut erinomaisesti. Sen sijaan suuri jäännöstermi kertoo mallin epäsopivuudesta ennustaa muuttujan arvo. Regressiomalli pyritään muodostamaan sellaiseksi, että kaikkien jäännöstermien neliöiden yhteenlaskettu summa saadaan mahdollisimman pieneksi. Tällöin kaikkien havaintopiteiden yhteenlaskettu etäisyys suoralla on mahdollisimman pieni. (Nummenmaa 2011, 310.)

Mallin tilastollinen merkitsevyys

Mallin tilastollinen merkitsevyystestaus on pätevä, mikäli lineaarisen regressiomallin edeltävyys ehdot täyttyvät: lineaarinen riippuvuus selittävien muuttujien ja selitettävän muuttujan välillä, kaikkien selittävien muuttujien arvoilla on yhtä suuri jäännösten varianssi sekä normaalijakautuneet ja toisistaan riippumattomat jäännökset (Taanila 2010, 11). Jäännöskuviossa piste x -akselilla kuvaa ennustettua arvoa ja y -akselilla havaintoa vastaavaa jäännöstermin arvoa. Jäännöstermin nollakohta on x -akselin suuntaisen viivan kohdalla. Toimivassa mallissa y -muuttujan pieniin ja suuriin havaittuihin arvoihin liittyy itseisarvoltaan pieniä ja suuria jäännöstermejä, ja pisteet jakautuvat satunnaisesti ja epäsäännönmukaisesti. (Taanila 2010, 13.) Jos edeltävyys ehdot täyttyivät riittävästi, koko mallin tilastollisen merkitsevyys voidaan testata F -testillä (Taanila 2010, 16).

Malli sopii keskimääräisen käyttäytymisen kuvaamiseen, jos jäännökset jakautuvat normaalisti (Saaranen 2017, 58). Normaalisti jakautuvan (normaalisen) kuvaajan jakauma on symmetrinen ja huippu osuu keskimääräiseen arvoon (Saaranen 2018, 18). Mikäli aineisto vastaa normaalijakaumaa ovat vinous- ja huipukkuusluvut arvoltaan 0. Positiiviset arvot viittaavat huipukkuuteen tai vinouteen oikealle ja negatiiviset arvot laakeaan jakaumaan ja vinouteen vasemmalle. Arvojen keskivirheen avulla voidaan päätellä niiden merkitsevyys. Jos keskivirheen arvo on < -2 tai > 2 , on se tilastollisesti merkitsevä. (Saaranen 2018, 19.)

Selittämättömän vaihtelun osuus ratkaisee mallin soveltuvuuden ennustamiseen. Sen suuruutta ja mallin sopivuutta havaintoaineistoon voi arvioida hajontakuviosta. Molemmiin

puolin mallia hajaantuvat havainnot kertovat ennusteeseen sisältyvästä epävarmuudesta. Koska mallit ovat herkkiä poikkeaville arvoille, tulee selvittää voiko kyseessä olla virheelliset tai väärin syötetyt tiedot tai löytyykö niille joku luonnollinen selitys. Virheelliset tiedot tulee korjata tai poistaa tarkastelusta tai jos luonnollinen selitys löytyy, tarkastella mallia ilmiön näkökulmasta. (Taanila 2015, 15.)

Aikasarjoja tarkasteltaessa tulee vielä huomioida ongelmallinen jäännösten mahdollinen riippuvuus, jota myös autokorrelaatioksi kutsutaan. Sen voi todeta hajontakuviosta: toisistaan riippumattomien jäännösten pistejoukossa ei ole satunnaisuudesta poikkeavaa säännönmukaisuutta. (Taanila 2010, 15.) Regressiomallin jäännöstermien korreloituminen leventää regressiokertoimien luottamusvälejä ja pienentää testisuureiden arvoja (Mellin 2006, 408). Durbin-Watsonin testillä voidaan osoittaa jäännösten riippumattomuus. Jos havainnot muuttuvat ajan kuluessa, ne ovat mitä ilmeisimmin toisistaan riippuvia eli esiintyy autokorrelaatio. Jos Durbin-Watsonin testin arvo on 1,5–2,5 ei jäännösten välillä ole autokorrelaatiota eli ne ovat riippumattomia. (Karadimitriou & Marshall 2020).

Ideaalinen regressioanalyysimalli on mahdollisimman yksikertainen, teoriaan pohjautuva, sopii aineistoon ja omaa mahdollisimman korkean selitysasteen (Nummenmaa 2011, 315–316). Lineaarisen mallin eli suoran avulla tutkitaan yhden selittävän muuttujan vaikutusta selitettävään muuttujaan. Mallina voidaan käyttää myös käyrää. (Saaranen 2018, 69.) Mallin sopivuutta voi kokeilla parantaa poistamalla pisteparven poikkeamat tai analysoimalla aineistoa epälineaarisella käyräsovituksella (Saaranen 2017, 60–61).

4.9 Tilastollinen analyysi

Määrällinen osuus tutkimuksesta toteutettiin SPSS-tilasto-ohjelmalla, jolloin aineistoja tilastollinen analyysi valittiin mukaan menetelmäksi. Tilastollinen analyysi jakautuu kuvailevaan analyysiin ja tilastolliseen päättelyyn. Kuvailevaan tilastoanalyysiin kuuluvat tunnusluvut, grafiikka, jakaumat ja riippuvuudet. (Saaranen 2018, 1.)

Kuvaileva analyysi

Tunnusluvut

Sijaintiluvut kuvaavat tyypillistä muuttujan arvojen suuruutta: esimerkiksi moodi useimmin esiintyvää havaintoarvoa, keskiarvo kuvaa suunnilleen havaintoarvojen suuruutta, mediaani keskimmäistä havaintoa ja fraktiilit havaintojen jakamista osiin (Nummenmaa 2011, 61, 63). Persentiili eli sadannes ilmaistaan muodossa $p\%$:n fraktiili, jolla tarkoitetaan muuttujan arvoa, jonka alle jää $p\%$ havainnoista. P on luku, joka sijoittuu välille 0–100. (Saaranen 2018, 27.)

Hajontaluvuilla kuvataan muuttujan arvojen jakautumista esimerkiksi keskiarvon ympärillä, jolloin niistä käytetään hajonta-termiä. Hajontalukuihin kuuluvat muun muassa vaihteluväli, keskihajonta, vinous ja huipukkuus. Vaihteluvälillä tarkoitetaan erotusta, joka saadaan, kun muuttujan suurimmasta arvosta vähennetään pienin arvo. Välin suuruus näyttää muuttujien arvojen vaihtelun. Keskihajonta kertoo havaintojen keskimääräisestä etäisyydestä jakauman keskiarvoon nähden. Vinossa jakaumassa suurin osa havainnoista on pienempiä tai suurempia kuin keskiarvo, ja huipukas jakauma on terävähuippuinen. (Nummenmaa 2011, 66–69, 71–72.)

Grafiikka

Graafiset esitystavat on valittu tapauskohtaisesti kullekin kuviolle, jotta ne olisivat mahdollisimman selkeitä ja ymmärrettäviä. Valinnassa on muun muassa huomioitu halutaanko arvoja vertailla keskenään vai tarkastella muuttujien välistä riippuvuutta.

Viivakuvion eli -kaavion arvopisteet kuvataan pisteellä ja yhdistetään viivalla toisiinsa (Taanila 2019a, 22). Viivakuviot sopivat aika- ja indeksisarjojen kuvaamiseen (Saaranen 2018, 46). Google Trends -palvelusta saatavat kuvaajat ovat yleisimmin viivakuvioita.

Pylväsdiagrammilla eli -kaaviolla kuvataan eri luokkiin kuuluvia havaintoja, joissa pylväiden korkeus kuvaa havaintojen määrää: mitä enemmän havaintoja, sitä korkeampi pylväs (Nummenmaa 2011, 76–77). Pystypylväät sopivat erityisesti määrällisestä muuttujasta muodostettujen luokkien kuvaamiseen, kun esitetään muun muassa lukumääriä, prosenttiosuuksia tai keskiarvoja (Taanila 2019, 18).

Laatikko–jana-kuvio soveltuu hyvin jatkuvan muuttujan jakauman esittämiseen, sillä siitä näkee sijainnin ja hajonnan. Myös muuttujien vertaaminen toisiinsa onnistuu helposti. Kuviossa puolet havainnosta sisältyvät laatikkoon, johon mediaani on merkitty poikkiviivalla. Laatikon yläosa kuvaa yläkvartaalia ja alaosa alakvartaalia. Janan päiden välissä ovat muut havainnot, paitsi voimakkaimmin poikkeavat havainnot, jotka on merkitty ympyrällä. Suurin osa havainnoista sisältyy siis laatikon ja janan sisälle. (Nummenmaa 2011, 83–84.)

Hajontakuviota eli -kaaviota soveltuu kahden jatkuvan muuttujan yhteisjakauman esittämiseen, jolloin selittävän muuttujan asteikko on x-akselilla ja selitettävän muuttujan y-akselilla. Suoran ympärillä esiintyvät pisteet osoittavat lineaarista riippuvuutta. (Saaranen 2018, 47.) Voidaan myös todeta, että pistejoukon säännönmukaisuus ilmaisee riippuvuutta (Taanila 2019a, 21).

Keskiarvon luottamusvälit esitetään SPSS-ohjelmalla kuviolla, jossa piste kuvaa keskiarvoa ja jana luottamusväliä (Saaranen 2018, 84).

Tilastollinen päättely

Tilastollisen päättelyn avulla arvioidaan otoksen tulosten paikkaansa pitävyyttä perusjoukossa. Toisin sanoen kuinka todennäköisesti tulokset voidaan yleistää perusjoukkoon. Otoksen tuloksista voidaan arvioida perusjoukon keskiarvoa luottamusväleillä sekä eroja ja riippuvuuksia tilastollisella testauksella. (Saaranen 2018, 76.) Tilastollisella testauksella selvitetään, onko otoksessa saatujen tulosten erojen taustalla sattuma vai ovatko ne yleistettävissä perusjoukkoon. Testaus etenee järjestelmällisesti vaiheesta toiseen: asetetaan hypoteesi, valitaan tilastollinen testi, valitaan merkitsevyystaso, suoritetaan testi ja tehdään johtopäätökset. (Saaranen 2018, 86.)

Keskiarvon virhemarginaali eli luottamusväli

Jos otoksen keskiarvo halutaan yleistää perusjoukkoon, tulee huomioida otantavirheen epävarmuus, jota mitataan virhemarginaalilla (Taanila 19.4.2019a). Luottamusväli on väli, jonka sisälle perusjoukon ei-tiedossa oleva keskiarvo osuu tietyllä luottamustasolla, joka tarkoittaa luottamusvälin todennäköisyyttä (Saaranen 2018, 80). Virhemarginaali kuvaa tiedon tarkkuutta, koska se perustuu otosten poiminnan satunnaisvaihteluun. Se ei kata tiedon keruun ja käsittelyn todellisia virheitä eli harhaa. (Saaranen 2018, 84.)

Otoskeskiarvon jakauma on normaalijakauman muotoinen, kun otoskoko on vähintään 30. Keskiarvon luottamusväli ilmaisee epävarmuutta, joka sisältyy keskiarvoestimaattiin. Pieni luottamusväli tarkoittaa populaatiokeskiarvon suuruuden luotettavuutta. Otoskeskiarvojen jakauman hajonta on keskiarvon keskivirhe. (Nummenmaa 2011, 142–143.) Tuloksia perusjoukkoon yleistettäessä esimerkiksi 95 % todennäköisyydellä otoksen keskiarvo on yleistettävissä saatuihin keskiarvojen raja-arvoihin. Toisin sanoen 5 % riskillä keskiarvo jää luottamusvälin ulkopuolelle. (Saaranen 10.3.2020.)

Merkitsevyystaso

Merkitsevyystaso eli riskitaso ilmaisee tuloksen tilastollisen merkitsevyyden tai yleistettäessä riippuvuus koko perusjoukkoon, virheen todennäköisyyden eli riskin sille, että havaittu riippuvuus tai ero johtuu sattumasta (Saaranen 2018, 64, 88).

Tuloksia tulkitaan seuraavasti:

- Suuntaa antava ($0,05 < \text{sig} < 0,10$)
- Tilastollisesti melkein merkitsevä ($0,01 < \text{sig} < 0,05$)
- Tilastollisesti merkitsevä ($0,001 < \text{sig} < 0,01$)
- Tilastollisesti erittäin merkitsevä ($\text{sig} < 0,001$).

4.10 Google Trends -hakukoneanalytiikkatyökalu

Google Trends -palvelun tarjoamat mahdollisuudet liiketoiminnan kehittämiseksi kiinnostivat kohdeyritystä, joten se valittiin mukaan tutkimukseen. Avoin työkalu sopii erinomaisesti tutkimuskäyttöön, sillä Google Trends hyödyntää Googlea, joka on maailman ja Suomen ylivoimaisesti suosituin hakukone. Lisäksi Google Trendsin tuottama data on massadataa, ja sillä voi tutkia hakujen yleisyyttä Suomessa ja maakunnittain. Hakujen ajoittumista voi tarkastella kellonajoittain, viikontähtäin, viikoin tai kuukausittain, vuosittain tai haluttuun ajanjaksoon. Vertailuja voi tehdä myös eri ajanjaksojen kuten vuosien välillä. Hakutermejä voi tarkentaa erilaisilla ehdoilla, joilla määritellään vaihtoehtoja (tai, ja) tai suljetaan pois tiettyjä sanoja (taulukko 2).

Taulukko 2. Google Trends -hakutermin tarkentaminen (Stephens-Davidowitz & Varian 2015)

Välimerkki	Merkitys	Esimerkki käyttämisestä
+	tai	[Lakers + Celtics]; hakutulokset sisältävät joko sanan [Lakers] tai sanan [Celtics]
-	jättää pois	[jobs – steve]; hakutulokset sisältävät [jobs], mutta eivät [steve]
tyhjä väli	ja	[Lakers Celtics]; hakutulokset sisältävät sekä sanan [Lakers] että sanan [Celtics], järjestys voi vaihdella
lainausmerkit	fraasi	[“Lakers Celtics”]; hakutulokset sisältävät eksaktisti Lakers Celtics

Google Trends on hakukoneanalytiikkatyökalu (Sponder & Khan 2018, 27), joka tarjoaa tietoa Google-hakukoneella tehdyistä kyselyistä hakutermin, kategorian ja maantieteellisen sijainnin perusteella. Hakujen tulokset esitetään indeksinä. Mitä suurempi indeksi on, sitä enemmän kyseistä hakutermiä on haettu. (Boone ym. 2015.) Indeksien maksimiarvo on 100. Esimerkiksi datapiste 50 tarkoittaa, että hakujen määrä on puolet pienempi verrattaessa sitä datapisteeseen 100 (Stephens-Davidowitz & Varian 2015).

Kategorioiden avulla voi selvittää hakua: esimerkiksi valitsemalle hakutermin [omena] kategoriaksi Tietokone & elektroniikka tai Ruoka & Juoma saa hakutermin haluamaansa kontekstiin (Stephens-Davidowitz & Varian 2015). Google kerää, luokittelee ja yhdistää hakudatan aiheeseen (Google 2020b). Google jakaa haut automaattisella luokittelulla yläkategorioihin ja edelleen toisen tason kategoriaan hyödyntämällä luonnollisen kielen käsittelyn metodeja. Esimerkiksi hakutermin [car tire] jaetaan kategoriaan Vehicle Tires, joka on Auto Parts -alakategoria, joka puolestaan on Automotive-alakategoria. (Choi & Varian 2009.)

On olemassa kahdenlaista satunnaisotoksena kerättävää Google Trends -dataa: reaaliaikaista, joka koostuu seitsemän edellisen päivän aikana tehdyistä hauista ja ei-reaaliaikaista, joka muodostuu datasta, joka on enintään vuodelta 2004 ja aikaisintaan 36 tuntia (1,5 vrk) vanhaa (Google 2020b). Jos haku ulottuu enintään kolme kuukautta taaksepäin, saadaan hakutulokseksi päivädataa, muutoin data on viikkodataa (Stephens-Davidowitz & Varian 2015).

Google Trends -datasta saadaan vertailukelpoista suhteuttamalla hakudata. Näin hakutermin suhteellista suosiota on mahdollista verrata. Hakutulokset suhteutetaan kyselyn ajankohtaan ja sijaintiin jakamalla jokainen datapiste alueen hakujen kokonaismäärällä ja ajanjaksolla. Jos suhteutusta ei tehtäisi, hakuintensiteetissä näkyisivät aina alueet, joilla haetaan eniten. Aiheen suosion mukaan, kaikkien hakuaiheiden ja hakujen joukossa, tulokset skaalataan asteikolle 0–100. Hakujen kokonaismäärä ei ole ehkä kuitenkaan sama, vaikka hakutermin suosio näyttäisi olevan sama eri alueilla. (Google 2020b.) Jos hakutermin indeksi pienenee ajan mittaan, ei tulos välttämättä tarkoita, että hakujen määrä on laskenut. Hakuja voi olla suhteellisesti entistä vähemmän prosentteina kaikista hauista. Kaikista aiheista tehdyt haut ovat lisääntyneet kuitenkin ajan myötä absoluuttisesti. (Stephens-Davidowitz & Varian 2015.) Toisaalta suurempi arvo merkitsee suhteellisesti isompaa osaa kaikista kyselyistä, mutta ei korkeampaa absoluuttista kyselyjen määrää. Käytännössä tämä tarkoittaa, että pienessä maassa 80 % hauista termillä [banaanit] saa tuplasti isomman tuloksen, kuin suuressa maassa, jossa hakutermi kattaa ainoastaan 40 % kyselyistä. (Google Trends 2020b.)

Google Trends -palvelun kartat kuvaavat hakujen intensiteettiä eri alueilla. Tummat sävyt kartalla kertovat alueet, joissa hakutermiä todennäköisimmin etsitään. Jos alueella ei ole väriä, ei hakutermistä ole riittävästi hakuja. (Google 2020b.) Google Trends -palvelun pylväskuvioista voi verrata alueiden välistä vaihtelua eri vuosien välillä.

Google Trendsin aikasarjojen avulla voi myös tutkia, kuinka jonkin hakutermin suosio on muuttunut ajan mittaan jollakin alueella. Esimerkiksi halutaan tietää, minä päivinä ihmiset etsivät eniten hakutermiä [martini recipe] ajanjaksolla marraskuu 2013–tammikuu 2014 USA:ssa. Hakutuloksista nähdään, että hakutermiä etsittiin eniten lauantaisin ja hakupiikki oli 31.12.2013 uudenvuodenaattona. Voidaan myös verrata kahta hakutermiä samalla ajanjaksolla ja alueella tai useita alueita samalla ajanjaksolla. Normalisaatio tehdään erikseen kaikille hauille, joten eri aikaan tehtyjen hakujen vertaaminen ei ole mielekäästä. (Stephens-Davidowitz & Varian 2015.)

Google Trendsillä ei voi saada hakujen tekijöistä yksilöivää tietoa esimerkiksi taustamuut-
tuvia kuten sukupuoli tai ikä, koska Google poistaa henkilötiedot. Google Trends ei analy-
soi harvinaisia hakuja, saman käyttäjän toistuvia hakuja lyhyen ajan sisällä ja erikoismerk-
kejä sisältäviä hakuja (esimerkiksi heitto- tai muut erikoismerkit). (Google 2020b.) Google
Trendsillä ei saa tietoja hakujen määristä.

4.11 Ubersuggest-hakukoneanalytiikkatyökalu

Tutkimukseen tarvittiin avoin hakukoneanalytiikkatyökalu, jotta päästiin analysoimaan ha-
kusanoja ja -volyymeja. Lisäksi sillä voi kerätä tietoa hakijoiden ikäryhmistä. Ubersuggest
nousi positiivisesti esille useissa ammattilaisten tekemissä työkaluarvioinneissa (Piippo
2016, Backlinko 2020). Lisäksi se on hyvä valinta Google Trendsiä täydentäväksi työka-
luksi, koska Ubersuggest kerää datan Googlesta käyttäen rajapintaa Google Adsiin
(Keyword Planner) (Bhat 29.2.2020).

Ubersuggest on suunniteltu toimimaan kuten Google Suggest, ja se hyödyntää samoja al-
goritmeja, joita hakukone ehdottaa avainsanoiksi omille käyttäjilleen. Tämä mahdollistaa
reaaliaikaiset avainsanaehdotukset ja tilastotiedot, jotka reagoivat automaattisesti Google-
hakuihin. Ubersuggest käyttää Googlen automaattisia ehdotuksia, kun liitetään yhteen va-
litun avainsanan kanssa etu- ja jälkiliitteitä. Mittarit voi poimia datasta analysoimalla yh-
teen liitettyjen avainsanojen rankkauksia. Sovelluksesta saa prosentteina tilastotiedot,
kuinka todennäköisesti avainsana tulee saamaan korkeamman rankkauksen. (Bynum
2019.)

Ubersuggest on monipuolinen työkalu, jolla voi tehdä hakusanatutkimuksen, kartoittaa
kuukausittaiset hakuvolyymit, katsastaa trendin vuositason sekä tutkia paluulinkkejä.
Ubersuggest-sovelluksessa voi domain-osoitteen antamalla tarkastella minkä vain verkko-
sivuston arvioitua orgaanista kuukausitrafiiikkia, rankkautumista eli sijoittumista Google-
hauissa (asteikko 1–100) ja muilta sivuilta verkko-osoitteeseen ohjaavien paluulinkkien
määrää. Orgaaninen kuukausittainen trafiiikki tarkoittaa verkkotunnuksen arvioitua koko-
naistrafiiikkia orgaanisilla eli ei-maksullisilla avainsanoilla. Verkkosivun vaikuttavuus, joka
määräytyy sivun pistemäärän perusteella, ratkaisee sijoittumiseen Google-hauissa. Arvo-
valtaisen sivun pistemäärä on suuri, ja se sijoittuu korkealle hakutuloksissa. (Ubersuggest
2019.) Ubersuggest on tehokas työkalu, joka auttaa pääsemään kärkeen Googlen rank-
kauksissa (Bynum 2019).

4.12 Excel-taulukkolaskentaohjelma

Tutkimuksessa hyödynnettiin taulukkolaskentaohjelmaa, jotta Google Trendsistä ladattuja csv-tiedostoja voitiin käsitellä helpommin, tallentaa xlsx-muotoon ja valmistella SPSS-ohjelman analyysijä varten. Microsoft Excel valittiin taulukkolaskentaohjelmaksi, koska siitä oli aikaisempaa kokemusta ja osaamista.

4.13 IBM SPSS Statistics -tilasto-ohjelma

Tutkimuksen tilastollisiin analyysihin käytettiin IBM SPSS-ohjelmistoa, josta löytyivät kaikki tarvittavat analyysit ja graafiset esitystavat.

IBM SPSS (Statistical Package for Social Sciences) Statistics -tilasto-ohjelma on tarkoitettu numeroaineistojen analysointiin. Ohjelmalla voi muun muassa käsitellä aikasarjadataa, tuottaa tilastollisia tunnuslukuja, korrelaatioita, erilaisia regressioanalyysijä ja ennusteita. (IBM 2020b.) SPSS on monipuolinen käyttäytymis- ja yhteiskuntatieteiden tarpeisiin suunniteltu ohjelma, joka taipuu melko monimutkaisiin analyysihin (Nummenmaa 2011, 21).

4.14 Tilastokeskuksen Tee oma karttaesitys -palvelu

Tutkimuksen spatiotemporaalista osuutta varten tarvittiin Suomen kartta maakunnittain, jossa voi esittää alueiden välisiä eroja väreillä. Karttapohja oli saatavana Tilastokeskuksen avoimesta Tee oma karttaesitys -palvelusta.

Palvelu perustuu Statistics eXplorer -verkkoyhteensopivaan sovellukseen. Tilastokeskuksen mukaan karttasovellus sopii paremmin suhteutettujen arvojen kuvaamiseen kuin absoluuttisten lukumäärien tutkimiseen informatiivisuuden kannalta, joten tutkimuksen viive-lukujen kuvaaminen kartoilla sopi näin ollen hyvin tarkoitukseen. (Tilastokeskus 2020a).

5 Tutkimuksen toteutus

Tutkimus toteutettiin soveltaen CRISP-DM-prosessia, koska se on erityisesti analytiikka-hankkeisiin kehitetty prosessimalli. Mallin valinnalla haettiin systemaattista lähestymistapaa tutkimukseen. Mallin eri vaiheet ja eteneminen on kuvattu kohdassa 1.6. Liiketoiminta- ja kuluttajaymmärrystä koskevat osuudet on koottu kohtaan 5.1., datan ymmärtäminen, kerääminen ja valmistelu kokonaisuus kohtaan 5.2. sekä analyysit ja mallinnus kohtaan 5.3., joka jakautuu edelleen kolmeen osaan: päivittäiselintarvikkeiden hakuajankohdat kohta 5.3.1., punaviinin Google-hakujen ja ostamisen välinen viive kohta 5.3.2. ja punaviinin myynnin ennustemalli kohta 5.3.3.

5.1 Liiketoiminta- ja kuluttajaymmärryksen parantaminen

Kohdeyrityksen liiketoimintaymmärryksen lisäämiseksi perehdyttiin yrityksen toimintaan ja toimialaan dokumenttiaineistoista, henkilökohtaisilla keskusteluilla toimeksiantajan ja kohdeyrityksen henkilöstön kanssa sekä kohdeyrityksen kehitysjohtajan puolistrukturoidulla haastattelulla. Tarkoituksena oli tunnistaa ratkaistava ongelma, joka tällä kehitysprojektilla on tarkoitus selvittää ja asettaa tavoitteet. Dokumenttiaineiston avulla tutkittiin lisäksi, milaista dataa data-analytiikan työkaluilla voi kerätä kuluttajaymmärryksen lisäämiseksi.

Puolistrukturoitu haastattelu

Puolistrukturoitua haastattelua varten mietittiin etukäteen teemat ja niihin liittyviä kysymyksiä. Haastattelulomake on salaisena liitteenä 2. Kohdeyrityksen kehitysjohtajaa haastatteli henkilökohtaisesti yhtä aikaa kaksi tutkimuksen tekijää kehittämistehtäviä varten. Haastattelussa pyrittiin lisäämään ymmärrystä tutkimuskohteesta ja esitettiin tarkentavia kysymyksiä. Haastattelu nauhoitettiin ja litteroitiin painottaen asioiden ymmärtämistä kokonaisuutena (salainen liite 3).

Dokumenttianalyysi

Dokumenttiaineistoja haettiin laajasti muun muassa Haaga-Helian kirjaston eri tietokannoista, Google Scholar-palvelusta, Google-hakukoneesta ja Iris.ai-tekoälypalvelun avulla. Lisäksi tiedonhakua syvennettiin sähköpostikyselyillä ja haastattelulla. Dokumenttiaineistona käytettiin monipuolisesti verkkosivuja, blogeja, artikkeleita, raportteja, tutkimuksia, kirjoja, tilastoja, haastattelua, henkilökohtaista tiedonantoa, videoita ja luentoja. Aineistoihin perehdyttiin ja valittiin tutkimuksen kannalta oleelliset sisällöt, jotka koottiin tekstitiedostoon ja tiivistettiin.

Liiketoiminta- ja kuluttajaymmärryksen parantamiseen liittyvät tulokset on koottu kohtiin 1.6.1., 3 ja 6.1.

5.2 Datan ymmärtäminen, kerääminen ja valmistelu

Kun liiketoiminnalliset tarpeet ja tutkimuksella selvitettävä ongelma oli selvitetty, hankittiin teoreettista ymmärrystä datasta ja selvitettiin saatavilla olevat datalähteet kohta 6.1. CRISP-DM-prosessin mukaan seuraavaksi edettiin datan ymmärtämiseen kohta 1.6.2. ja datan valmisteluun kohta 1.6.3. Kartoitettiin tutkimuksessa tarvittava data. Päivittäiselintarvikkeiden hakuajankohtien selvittämiseen ja muihin analyyseihin tarvittavat hakudatat saatiin ladattua Google Trends -palvelusta. Google-hakujen ja ostamisen välisen viiveen selvittämiseen sekä myynnin ennustemalliin tarvittiin lisäksi avointa päivä- ja tuntidataa soveltuvan päivittäiselintarvikkeen myynneistä. Toisen osapuolen avointa dataa oli saatavana Alkolta alkoholijuomien myyntitilastoista Suomesta ja maakunnista, mutta vain kuukausitasolla. Alkoholiksi sopi tutkimuskohteeksi, koska alkoholin vähittäismyynti on osa päivittäistavaramarkkinoita EU:ssa (Päivittäistavarakauppa ry 2020). Suomessa alkoholipitoisuudeltaan alle 5,5 % viinejä voi ostaa päivittäistavara-kaupoista, mutta sitä väkevämmät viinit on ostettava Alkosta (Baraka 2019). Tutkittavaksi kohteeksi valittiin punaviini, koska se nousi sesonkituotteiden kartoituksessa suositukseksi joulutuotteeksi (liite 2), josta oli olemassa myös jonkin verran taustatietoa. Punaviinin kuukausikohtaista myyntidataa ei saatu Alkolta digitaalisessa muodossa, joten se poimittiin käsin pdf-tiedostoista Alkon verkkosivuilta (Alko 2019).

5.2.1 Päivittäiselintarvikkeiden hakuajankohdat

Kartoitettiin aluksi tiedostoista, kuinka tarkkaa dataa eri aikajännteillä Google Trendsistä on saatavana (taulukko 3). Esimerkiksi seitsemän päivän aikajännteellä päästään tuntien tarkkuuteen, mutta viiden vuoden ajalla vain päivän tarkkuuteen, josta saadaan Excelin funktiotoiminnolla kuukaudet ja viikkonumerot esille, mutta ei viikonpäiviä. (Google 2020b.)

Taulukko 3. Google Trends -palvelusta saatavan datan tarkkuus (Google 2020b)

	Pvm	Kuukausi ¹⁾	Viikkonumero ¹⁾	Viikonpäivä ¹⁾	Tunti
Viimeiset 5 vuotta	x	x	x	-	-
Viimeiset 90 päivää	x	x	x	x	-
Viimeiset 7 päivää	x	x	x	x	x

¹⁾ Saatavana Google Trendsin datan päivämäärästä Excelin funktiotoiminnolla

Valittujen päivittäiselintarvikkeiden hakuajankohtien selvittämiseen käytettiin Google Trends -palvelusta saatavaa dataa. Kaikki haut rajoitettiin Suomeen. Datat ladattiin palvelusta csv-muodossa. Datat vietiin Exceliin, jossa tiedot siirrettiin sarakkeisiin, muokattiin päivämäärät ja tallennettiin xlsx-muotoon. Excelin funktiotoinnolla taulukkoon lisättiin sarakkeet kuukaudelle sekä viikon- ja viikonpäivänumerolle. Pohjatietoina käytettiin datan alkuperäisiä päivämääriä. Kaikki tiedot tarkastettiin vertaamalla kalenteritietoihin (Helsingin yliopiston kalenteripalvelu 2014, 2017 & 2018; Kalenteripalvelu Oy 2015 & 2016). Viikonnumerot eivät vastanneet kalenteritietoja, ja ne korjattiin vähentämällä 1–2.

Kuukausi- ja viikkokohtaiset erot: 5 vuoden ja 90 päivän datat

Google Trends -datat käsiteltiin muotoon, jossa SPSS-ohjelmalla voitiin muodostaa tuotteen viiden vuoden hakuindeksien keskiarvojen määrän jakaumista kuukausittain laattikko–jana-kuviot ja viikoittain pylväsdiagrammit. Uusien kasviproteiinien kohdalla oli kaksi poikkeusta tarkasteluajanjaksoissa. Nyhtökaura tuotiin markkinoille toukokuussa 2016, ja haut alkoivat vasta 20.12.2015, joten datat poistettiin siihen asti viikkotarkastelussa. Härkis puolestaan lanseerattiin syyskuussa 2016. Sitä alettiin hakea 21.8.2016 alkaen, jota edeltävät datat poistettiin.

Viikonpäivien ja kellonaikojen erot: 90 ja 7 päivän datat

Viideltä vuodelta kerätyssä Google Trends -datassa kaikki viikonpäivät olivat sunnuntaita, joten viikonpäivien välisen eron selville saamiseksi tarvittiin tarkempaa dataa. Kerätiin tuotteen hakuindeksien 90 päivän data, johon lisättiin Excelin funktiotoinnolla viikonpäivät. Ei-sesonkituotteiden kellonajat kartoitettiin 22.1.2020 seitsemän edellisen päivän hakuindeksien keskiarvoista. Selvitettiin 25.12.2019 joulusesonkituotteiden seitsemän edellisen päivän hakuindekseistä hakuajankohdat päivittäin ja kellonajoittain viikon aikajännteellä.

Google Trendistä kerättyjen ja Excelissä käsiteltyjen datojen mallitaulukot on koottu liitteeseen 3.

5.2.2 Punaviinin Google-hakujen ja ostamisen välinen viive

Google-hakujen ja ostamisen välisen viiveen selvittämiseksi Suomessa kerättiin datat ristikorrelaatioanalyysin toteuttamista varten. Google Trends -haut tehtiin sanalla [punaviini] ajanjaksolla 1.9.2016–30.10.2019. Google Trends -asetuksissa määritettiin alueeksi Suomi, [punaviini] tarkennettiin alkoholijuomaksi ja kategoriaksi valittiin Ruoka ja juoma.

Google Trendsin kuukausitason data sisältää 165 havaintoa. Alkon myyntitilastoista kerättiin punaviinin kuukausikohtaiset myyntimäärät litroina Suomessa syyskuusta 2016 lokakuuhun 2019. Datasetti sisältää 38 kuukauden myyntitiedot.

Google Trends -haut tehtiin sanalla [punaviini] ajanjaksolla 1.9.2016–30.10.2019 maakunnissa spatiotemporaalista analyysia varten. Maakuntakohtaiset haut tehtiin vastaavilla asetuksilla kuin koko Suomea koskevat haut edellä, mutta alialueesta valittiin kukin kahdeksastatoista maakunnasta. Maakunnat ovat Uusimaa, Varsinais-Suomi, Satakunta, Kanta-Häme, Pirkanmaa, Päijät-Häme, Kymenlaakso, Etelä-Karjala, Etelä-Savo, Pohjois-Savo, Pohjois-Karjala, Keski-Suomi, Etelä-Pohjanmaa, Pohjanmaa, Keski-Pohjanmaa, Pohjois-Pohjanmaa, Kainuu ja Lappi. Alkon myyntitilastoista kerättiin punaviinin kuukausikohtaiset myyntimäärät litroina 18 eri maakunnassa (pois lukien Ahvenanmaa) syyskuusta 2016 lokakuuhun 2019. Myyntidatasta poistettiin Ahvenanmaa, koska Google Trendsin alialueissa ei ole saatavana Ahvenanmaata koskevaa tietoa. Datasetti sisältää 38 kuukauden myyntitiedot.

Datasettien yhdistäminen

Koko Suomen ja maakuntien Alkon myyntidatat ja Google Trends -hakujen indeksit yhdistettiin Excel-taulukoksi (liite 3). Google Trends -indekseissä on kunkin kuukauden kohdalla 4–5 päivämäärää vastaavaa indeksiä, yhteensä 165 datapistettä. Alkon tilastoissa punaviinin myynnille on yksi kuukausikohtainen luku, yhteensä 38 kuukautta. Kuukausittaiset Alkon myyntimäärät kopioitiin sellaisenaan kunkin Google Trends -indeksin kohdalle, jolloin rivejä tuli yhteensä 165, mikä vastaa indeksin datapisteiden määrää.

5.2.3 Punaviinin myynnin ennustemalli

Alkon myyntitilastoista koottiin punaviinin kuukausikohtaiset myyntimäärät litroina Suomessa syyskuusta 2016 lokakuuhun 2019 Excel-taulukoksi. Datasetti sisältää 38 kuukauden myyntitiedot.

Google Trends -haut tehtiin sanalla [punaviini] ajanjaksolla 1.9.2016–30.10.2019 kuukausittain Suomessa. Google Trends -asetuksissa määritettiin alueeksi Suomi, [punaviini] tarkennettiin alkoholijuomaksi ja kategoriaksi valittiin Ruoka ja juoma. Indeksissä on 165 datapistettä. Datat ladattiin Google Trends -palvelusta Google-hakuja kuvaavina indekseinä csv-tiedostoina.

Datasettien yhdistäminen

Käytettiin samaa Excel-yhdistelmätaulukkoa kuin edellä (kohta 5.2.2.), josta hyödynnettiin Suomea koskevat Alkon myyntidatat ja Google Trends -hakujen indeksit (liite 3).

Myynti- ja indeksidatan tunnusluvut

Tehtiin tunnuslukuyhteenvedot koko Suomen punaviinin Google Trends -indeksi- ja myyntidatasta SPSS-ohjelmistolla. Asetukset on kuvattu liitteessä 4.

5.3 Analyysit ja mallinnus

Datojen valmistelun jälkeen edettiin analysointivaiheeseen. Tässä osuudessa kuvataan, kuinka analyysit on tehty ja miten mallit on rakennettu. Kokonaisuus jakautuu kolmeen osaan: päivittäiselintarvikkeiden hakuajankohdat kohta 5.3.1., punaviinin Google-hakujen ja ostamisen välinen viive kohta 5.3.2. ja punaviinin myynnin ennustemalli kohta 5.3.3.

5.3.1 Päivittäiselintarvikkeiden hakuajankohdat

Tutkittiin päivittäiselintarvikeryhmistä meijerituotteiden, valmisruokien, kasviproteiinien ja sesonkituotteiden tiedonhakemisen ajankohtaa Googlesta. Lisäksi tutkittiin punaviinin hakuja. Google Trends -indeksien avulla haluttiin selvittää, milloin ihmiset alkavat valmistautua oston ja minä viikonpäivinä ja kellonaikoina jotakin tiettyä tuotetta ajatellaan ja haetaan Googlesta. Ensimmäinen ajankohta on tärkeä tieto massamedian myynnin ja jälkimäinen mainoksen oikean ajoittamisen kannalta.

Elintarvikeryhmien tuotenimikkeiden määrittäminen

Elintarviketuoteryhmien tuotenimikkeiden määrittämisessä käytettiin ensisijaisesti Tilastokeskuksen yksilöllisen kulutuksen käyttötarkoituksen mukaista COICOP-luokitusta. Maitotuotteiden ja valmisruokien tuoteryhmät pilkottiin tuotetasolle COICOP-luokituksella. Valmisruoista tarkasteltiin vain aterioita tuoteryhmän laajuuden takia.

Hakutrendien analysointi

Aluksi selvitettiin millä sanoilla kuluttajat hakevat tutkittavia elintarvikkeita. Tarkasteltiin kerättyjen tuotenimikkeiden hakusanoja ja -volyymeja Übersuggest-palvelussa. Katsottiin tuotenimien hakuvolyymit ja luokituksen nimien rinnakkaiset hakusanat esimerkiksi erilaiset kirjoitusasut, jotka nousivat esille. Valittiin analyysin perusteella tuotteet ja käytettävät hakutermit Google Trends -hakuja varten. Hakusanoissa huomioitiin myös poissuljettavat sanat. Esimerkiksi meijerituotteita haettaessa nousi esiin, että esim. juustoa haetaan myös hakutermeillä maidoton, vege ja vegaaninen, jotka suljettiin pois hauista, koska haluttiin tutkia maitoperäisiä tuotteita.

Yhteenvedotaulukot elintarvikeryhmien tuotenimikkeiden sekä hakusanojen ja -volyymien määrittämisestä ovat liitteenä 5.

Meijerituotteet

COICOP-luokituksen mukaisilla sanoilla saatiin meijeri- eli maitotuotteille Ubersuggest-palvelusta hakuvolyymit (28.11.2019), valittiin Google Trends -hakusanat, ja tiedot koottiin taulukkoon (lyhennelmä taulukossa 4, koko taulukko liitteenä 5). Valittiin kolme suosituinta tuotetta Ubersuggest-hakuvolyymien perusteella: voi (5 400), jäätelö (4 400) ja juusto (2 400) Google Trends -hakuja varten. Asetettiin hakuihin suodattimet ja rajaukset Ubersuggest hakujen perusteella. Esimerkiksi huomattiin, että koska halutaan tutkia nimenomaan meijerituotteita (maitotuotteita) tulee kasvisperäiset maitoa korvaavat tuotteet rajata pois hausta. Meijerituotteiden Google Trends -hakutermeinä käytettiin [voi -maidoton], [jäätelö + kermajäätelö + maitojäätelö -kasvis -maidoton -vege -vegaani] ja [juusto -vege -maidoton -kasvis]. Luokaksi valittiin kaikissa Ruoka ja juoma.

Taulukko 4. Meijerituotteiden hakusanat COICOP-luokituksen (Tilastokeskus 2019b) mukaan, hakumäärät (Ubersuggest) ja valitut Google Trends -hakusanat

COICOP-luokitus	Hakusanat	Hakumäärät/ -volyymi	Valitut Google Trends -hakusanat
01.1.4.4 Juustot ja rahkat	juustot juusto	2400 2400	juusto
01.1.5.1 Voi	voi	5400	voi
01.1.8.5.1 Jäätelöt	jäätelöt jäätelö kermajäätelö maitojäätelö	4400 4400 90 90	jäätelö kermajäätelö maitojäätelö

Valmisruoat

Valmisruokien ateriar ryhmän COICOP-luokituksen mukaiset hakusanat analysoitiin ja katsottiin hakuvolyymit Ubersuggest-palvelussa 28.11.2019, jonka perusteella valittiin hakusanat Google Trends -hakuja varten ja tiedot koottiin taulukkoon (lyhennelmä taulukossa 5, koko taulukko liitteenä 5).

Valittiin Google Trends -hakuihin kolme tuotetta Ubersuggest-hakusana-analyysin perusteella (taulukko 6): maksalaatikko (5 400), koska se on Suomen ostetuin valmisruoka-ateria lähes 9 miljoonalla myydyllä rasiolla per vuosi (Paljakka 2019) sekä kinkkukiusaus (22 200) ja kasvissosekeitto (14 800) suurien hakumäärien takia. Lisäksi kaikista tuotteista löytyi hakuja yhdistettynä valmistajaan. Haettiin 14.2.2020 Google Trendistä hakutermillä [maksalaatikko Saarioinen + maksalaatikko Atria + maksalaatikko Pirkka + maksalaatikko

Rainbow], [kinkkukiusaus Saarioinen + kinkkukiusaus Kokkikartano + kinkkukiusaus Wotkins + kinkkukiusaus Pirkka] ja [kasvissosekeitto Rainbow + kasvissosekeitto Pirkka + kasvissosekeitto Apetit + kasvissosekeitto Saarioinen]. Luokaksi valittiin Ruoka ja juoma kaikkiin hakuihin. Google Trends -haussa oli alun perin mukana myös [maksalaatikko + Kokkikartano], mutta se jouduttiin jättämään pois virheilmoituksen takia: ”Sivua näytettäessä tapahtui virhe.”

Taulukko 5. Valmisaterioiden hakusanat COICOP-luokituksen (Tilastokeskus 2019c) mukaan, hakusanat ja -määrät (Ubersuggest) ja valitut hakusanat Google Trends -hakuja varten

COICOP-luokitus, ateriat	Hakusanat	Hakumäärät/-volyymit	Valitut Google Trends -hakusanat
Laatikat, keitot, salaatit, pastaruokat			
M0111102 Maksalaatikko	maksalaatikko maksalaatikko Saarioinen maksalaatikko Atria maksalaatikko Pirkka	5400 210 20 10	maksalaatikko
M0112704 Lihakaali-, lihaperunase- ym. laadikat	kinkkukiusaus Saarioinen Kokkikartano HK Pirkka	22200 - - 0 70	kinkkukiusaus
M0117607 Kasviskeitot, -laadikat ja -eineket	kasvissosekeitto Atria Rainbow Pirkka Apetit	14800 0 10 20 30	kasvissosekeitto

Kasviproteiinit

Kasviproteiinien tuotenimikkeiden hakusanat analysoitiin Ubersuggest-palvelussa 28.11.2019, kartoitettiin hakumäärät, valittiin hakusanat Google Trends -hakuja varten ja tiedot koottiin taulukkoon (lyhennelmä taulukossa 6, koko taulukko liitteenä 5).

Taulukko 6. Kasviproteiinien luokittelu (vegaanituotteet.net 2019), hakusanat ja hakumäärät (Ubersuggest) ja Google Trends -hakuihin valitut hakusanat (koko taulukko liitteenä 5)

Luokittelu	Hakusanat	Hakumäärät/-volyymit	Valitut Google Trends -hakusanat
Lihan sijaan käytettävät	Nyhtökaura Härkis soija rouhe soijarouhe	6600 3600 3600 2900	Nyhtökaura Härkis soija rouhe soijarouhe

Valittiin kolme suosituinta kasviproteiinia Ubersuggest-hakuvolyymien perusteella: Nyhtökaura (6 600), soija rouhe/soijarouhe (3 600/2 900) ja Härkis (3 600). Kaikki tuotteet ovat ruoanlaiton raaka-aineita, joten hakuja ei rajattu mitenkään. Huomattiin, että soijarouhetta haettiin runsaasti myös kirjoitusasulla ”soija rouhe”, joten molemmat kirjoitusasut lisättiin

hakutermiin. Google Trends -hakutermeinä käytettiin [Nyhtökaura], [soija rouhe + soija-rouhe] ja [Härkis]. Luokaksi valittiin kaikissa Ruoka ja juoma.

Sesonkituotteet

Joulusesongin tärkeimpien tuotteiden hakusanoihin liittyvät hakumäärät kartoitettiin, ja katsottiin hakupiikkien ajankohdat ja suuruudet Ubersuggest-palvelussa 28.11.2019 ja tiedot koottiin taulukkoon (lyhennelmä taulukossa 7, koko taulukko liitteenä 5). Valittiin joulusesongin kolme suosittua tuotetta Ubersuggestin joulukuun hakuvolyymien perusteella Google Trends -hakuja varten. Porkkanalaatikon (165 000) ja lanttulaatikon (135 000) hakuvolyymit olivat suurimmat, ja joulukinkku (hakuvolyymi 40 500) valittiin, koska siitä oli saatavana tietoa myyntiajankohdista. Hakutermejä ei rajattu, koska hakua tekevä ajattelee tuotetta juuri silloin, ja mahdollisesti harkitsee, ostaako tuotteen valmiina vai tekeekö itse. Esimerkiksi joulukinkun reseptiä tai paisto-ohjetta hakeva tarvitsee myös itse kinkun. Joulusesonkituotteiden Google Trends -hakutermeinä käytettiin [porkkanalaatikko]/Ruoka, [lanttulaatikko]/Laatikkoruoka ja [joulukinkku]/Ruoka. Luokaksi valittiin kaikissa Ruoka ja juoma.

Taulukko 7. Joulusesongin tärkeimmät tuotteet (Tilastokeskus 2018a), hakumäärät ja -piikit sekä yleisimmät hakuajankohdat (Ubersuggest) (koko taulukko liitteenä 5)

Hakusanat	Hakumäärät/ -volyymit	Hakuhiiput	Hakuajankohdat
joulukinkku	4400	40500	joulukuu
lanttulaatikko	12100	135000	joulukuu
porkkanalaatikko	18100	165000	joulukuu

Punaviini

Katsottiin Ubersuggest-palvelussa 15.1.2020, kuinka suuria punaviinin hakuvolyymit ovat eri kuukausina ja millaisia hakusanaideoita punaviinistä nousee esille (liite 6). Ubersuggestin mukaan punaviinin hakuvolyymi on 1 600 ([punaviini]-sanon hakumäärä kuukaudessa) Suomessa. Haussa eivät ole mukana esimerkiksi punaviinikastike (5 400), punaviini kalorit (720) tai punaviinimarja (720).

Google Trends -analyysit

Tuotteen trendi ja kausiluonteisuus: 5 vuoden aikajänne

Google Trendsin tuotteen hakuindeksejä kuvaavista viivadiagrammeista arvioitiin pitkän ajan (viiden vuoden) trendi, joka voi olla nouseva, stabiili tai laskeva. Tuotteen vuosittainen vaihtelu eli kausiluonteisuus selvitettiin hakemalla tuotetta viitenä eri vuotena 2015–2019. Mikäli eri vuosien hakuindeksejä kuvaavissa viivadiagrammeissa näkyy selkeä

huippu samaan aikaan eri vuosina ja muina aikoina hakuja on vain vähän, on kyseessä kausiluonteinen tuote, jota kutsutaan jatkossa myös sesonkituotteeksi.

Kuukausi- ja viikkokohtaiset erot

SPSS-ohjelmalla selvitettiin viiden vuoden keskiarvoista tuotteen Google Trends -hakuindeksien kuukausikohtaiset erot laatikko-janakuviolla sekä viikonnumeroiden väliset erot pylväsdiagrammilla.

Viikonpäivien ja kellonaikojen erot

Tehtiin SPSS-ohjelmalla viikonpäivien keskiarvoja osoittavat pylväsdiagrammit. Käytettiin samaa menetelmää ei-sesonki- ja sesonkituotteiden kohdalla, vaikka sesonkituotteissa ei hakuja aina ollut kolmen kuukauden ajanjaksolla lyhyestä hakuintensiteetistä johtuen. Tämä ei kuitenkaan heikentänyt keskiarvona esitettäviä tuloksia. Google Trendsin viimeisen seitsemän päivän hakuindeksien dataa käytettiin kuvaamaan viikon sisäisiä hakuvaihteluita viivadiagrammina.

Eri viikonpäivien kellonaikojen erot visualisointiin pylväsdiagrammeiksi SPSS-ohjelmalla. Ei-sesonkituotteiden hakujen kellonajat kartoitettiin 22.1.2020 seitsemän edellisen päivän Google Trends -hakuindeksien keskiarvoista. Joulusesonkituotteilta, joilla on selkeä hakuhiippu, selvitettiin seitsemän edellisen päivän hakuindekseistä 25.12.2019 ha-kuajankohdat päivittäin ja kellonajoittain viikon aikajänteellä.

Keskiarvon luottamusväli

Kuukausien, viikkonumeroiden, viikonpäivien ja kellonaikojen keskiarvoon perustuvien Google Trends -indeksien tulosten luotettavuus testattiin keskiarvon 95 %:n luottamusväleillä SPSS-ohjelmalla.

SPSS:llä tehtyjen tarkastelujen asetukset on kuvattu liitteessä 4. Tulokset ovat kohdassa 6.2.

Alialueet: maakunnat Suomen kartalla

Google Trendistä kerättiin tuotteiden hakujen yhteydessä myös hakujen intensiteettiä kuvaavat Suomen kartat ja alialueista maakuntakohtaisia hakuja kuvaavat pylväskuviot vuosille 2015–2019. Tuotekohtaiset tulokset ovat liitteinä 12–22.

Hakijoiden ikäryhmät

Ubersuggest-palvelusta saatiin lisäksi tietoa tuotteen hakijoiden iästä. Hakijat on luokiteltu palvelussa seuraaviin ikäryhmiin: < 18, 18–24, 25–34, 35–44, 45–54, 55–64 ja 65+. Haut

tehtiin palvelussa 25.1.2020. Yhteenveto tuloksista on liitteenä 7, ja tuotekohtaiset tulokset löytyvät liitteistä 12–22.

5.3.2 Punaviinin Google-hakujen ja ostamisen välinen viive

Google Trends -hakuindeksien ja punaviinin myynnin välistä viivettä tutkittiin kahden aikasarjajamuuttujan ristikorrelaatioanalyysillä keskiarvoista. Muuttujat ovat Google Trends -indeksi ja punaviinin myynti. Punaviinin myynnin ja hakemisen ristikorrelaatioanalyysi Suomessa ja maakunnissa tehtiin SPSS-ohjelmalla. Koska maakuntia koskeva aineisto on hyvin laaja, päätettiin rajata Suomea koskevien tulosten perusteella viive kolmeen luokkaan 0–2, jotka olivat tilastollisesti merkitseviä. Maakuntien ristikorrelaatiotaulukoista (liite 8) koottiin viiveluvut 0, 1 ja 2 ja kuvaajien 95 %:n luottamusväleistä tilastolliset merkitsevyydet Excel-taulukoon (liite 9). Merkitsevät arvot merkittiin harmaalla pohjavärillä taulukkoon. Dataan lisättiin viralliset kolminumeroiset maakuntakoodit (Tilastokeskus 2019g), ja maakunnat järjestettiin numerojärjestykseen pienimmästä suurimpaan.

Maakuntien Google-hakujen ja myyntien välisiä eroja havainnollistettiin tekemällä kartat Excel-taulukon datasta (liite 9) Tilastokeskuksen Tee oma karttaesitys -palvelulla. Negatiivisten korrelaatiokertoimien arvot aiheuttivat ei-toivotun poikkeaman karttojen väreihin. Asia ratkaistiin tekemällä luokitteludata Excel-taulukkoon (liite 10), jossa tilastollisesti merkitsevät kohtalaisen positiivisen riippuvuuden (0,3–0,7) omaavat lukuarvot merkittiin numerolla 2 (punainen väri kartalla) ja ei-riippuvuutta (-0,3–0,3) osoittavat arvot numerolla 1 (sininen väri kartalla). Numero 3 kuvaa arvoja, jotka eivät ole tilastollisesti merkitseviä (valkoinen väri kartalla). Kartat muodostettiin uudelleen.

SPSS-ohjelman ja Tee oma karttaesitys -palvelun asetukset on kuvattu liitteessä 4. Tulokset on esitetty kohdassa 6.3.

5.3.3 Punaviinin myynnin ennustemalli

Aluksi tutkittiin muuttujien välistä riippuvuutta. Tarkastelujen perusteella valittiin tutkimusmenetelmiksi Pearsonin korrelaatio ja regressioanalyysi. Ne sopivat menetelmiksi, koska toisiinsa kytkeytyvät korrelaatio- ja regressioanalyysit tuottavat tunnuslukuja aineiston rakenteesta ja tarjoavat työkaluja ennustamiseen. Ensin korrelaatioanalyysillä tarkastellaan muuttujien riippuvuuksia. Mallinnukseen käytetään regressioanalyysiä, joka on korrelaatioanalyysiä tarkempi menetelmä, jolla saadaan esille riippuvuuden matemaattinen muoto. (Kananen & Makkonen (toim.) 2015, 314.)

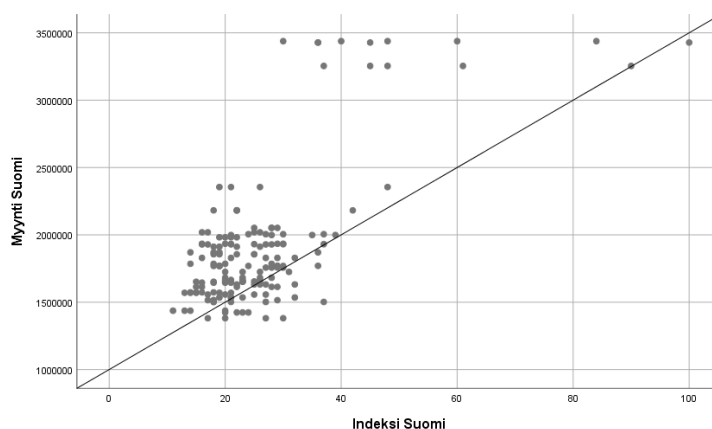
Google Trends -indeksi valittiin selittäväksi muuttujaksi, koska kohdeyritys oli kiinnostunut Google Trends -työkalun tuottaman yrityksen ulkoisen massadatan hyödyntämismahdollisuuksista omassa liiketoiminnassaan. Myynti valittiin selitettäväksi muuttujaksi, koska haluttiin ennustaa myyntimääriä. Lähdettiin mallintamisessa liikkeelle yksinkertaisella yhden selittävän muuttujan mallilla, koska näin voitiin tutkia puhtaasti Google Trends -indeksien vaikutusta myyntiin. Jos selittäviä muuttujia on useita, ne voivat korreloida keskenään, jolloin ne saattavat selittää samaa vaihtelua. Tällöin ei tiedetä, miten kahden muuttujan selittävä vaihtelu jakautuu niiden kertoimiin. (Taanila 2010, 7.)

Kahden muuttujan välisen riippuvuuden tutkiminen

Järkevyydestä tehtiin tutkimuksen muuttujille eli punaviinin myynnille (litraa) ja hakuintensiteettiä kuvaavalle Google Trends -indeksille (prosenttiosuus). Boonen ym. (2015) ovat tutkineet elintarvikkeiden Google Trends -indeksien ja myyntien korrelaatiota, joten riippuvuuden tarkastelu vaikuttaa perustellulta.

Koska sekä selittävä (indeksi) että selitettävä muuttuja (myynti) ovat välimatka- tai suhteasteikollisia voidaan valita analysointitavaksi Pearsonin korrelaatio, hajontakuviot, selityste sekä regressioanalyysi, mikäli muuttujien välillä ilmenee korrelaatiota. Molemmat muuttujat ovat jatkuvia, joten Pearsonin korrelaatiokerroin sopii kuvaamaan riippuvuuden voimakkuutta myös tältä näkökulmalta.

Tarkasteltiin vielä muuttujien välisen riippuvuuden määrää ja laatua hajontakuviolla (kuvio 7), joka toteutettiin SPSS-ohjelmalla. Y on selitettävä muuttuja eli punaviinin myynti ja x on selittävä muuttuja eli Google Trends -indeksi. Hajontakuvion pisteet seuraavat suoran linjaa nousemalla oikealle. Voidaan katsoa, että indeksin ja myynnin välillä on lineaarinen ja positiivinen riippuvuus (Saaranen 2017, 46).



Kuvio 7. SPSS:n hajontakuviot punaviinin myyntien ja hakuindeksien riippuvuudesta

Pearsonin korrelaatiomatriisin toteuttaminen

Kahden muuttujan, Google Trends -indeksien ja punaviinin myynnin, välistä lineaarista riippuvuutta Suomessa tutkittiin Pearsonin korrelaatiolla. Muodostettiin korrelaatiomatriisi SPSS-ohjelmalla (liite 11). Poimittiin matriisista indeksiin ja myynnin risteämiskohdasta korrelaatiokerroin ja p-arvo (kohta 6.4. tulokset).

Lineaarisen regressioanalyysin toteuttaminen

Regressioanalyysillä ennustettiin Google Trends -hakuindeksien vaikutus punaviinin myyntimääriin Suomessa. SPSS-ohjelmalla toteutettiin lineaarinen regressioanalyysi ja hajontakuviot sekä laskettiin ennusteiden keskiarvojen luottamusvälit havaintojen perusteella. Lisäksi arvioitiin mallin ennusteiden tarkkuutta havainnoista lasketuista ennusteiden luottamusväleistä. SPSS Statistics Data Editorista tiedot tallennettiin Excel-tiedostoksi ja tehtiin tunnuslukuyhteenvedot virhetermeille sarakkeesta RES_1 SPSS-ohjelmistolla. Jäännöstermien jakaumalla tutkittiin punaviinin myynnin osaa, jota ei voi ennustaa mallilla. Ominaisuuksien tilastollinen merkitsevyys testattiin SPSS-ohjelmistolla.

Mallin tilastollinen merkitsevyys

Mallin tilastollinen merkitsevyys selvitettiin edeltävyysehtojen testauksella, joka tehtiin SPSS-ohjelmalla. Aluksi laadittiin hajontakuviot, ja tarkasteltiin lineaarisuutta ja jäännösten variansseja, joiden tulisi olla yhtä suuria. Seuraavaksi tehtiin kuviot jäännösten normaalijakautuneisuuden tarkasteluun. Lopuksi jäännösten riippumattomuus testattiin Durbin-Watsonin-testillä.

Mallin parantaminen epälineaarisella käyräsovituksella

Koska lineaarisella regressiomallilla vajaan puolet punaviinin myynnin vaihtelusta Suomessa voitiin selittää hakuintensiteetin vaihtelulla (tulokset kohta 6.4), haluttiin testata, voiko mallin sopivuutta parantaa epälineaarisella käyräsovituksella. Laskelmat tehtiin SPSS-ohjelmalla. Paras malli löydettiin vertailemalla selityssasteita, joista valittiin suurin arvo, ja muodostettiin mallia kuvaava käyrä.

SPSS-ohjelmalla tehtyjen analyysien asetukset on kuvattu liitteessä 4. Tulokset on esitetty kohdassa 6.4.

6 Tutkimuksen tulokset

Kuluttajaymmärrystä data-analytiikalla osuuden dokumenttianalyysin tulokset löytyvät kohdasta 6.1. Päivittäiselintarvikkeiden hakuajankohtia koskevat tulokset on koottu kohtaan 6.2. Kohdassa 6.3. esitellään punaviinin Google-hakujen ja ostamisen välisestä viiveestä kertovat löydökset ja viimeisessä kohdassa 6.4. ovat punaviinin myynnin ennustemallin tulokset.

6.1 Kuluttajaymmärrystä data-analytiikalla

Datan määrä lisääntyy koko ajan kiihtyvällä vauhdilla, ja yrityksen tulee miettiä omassa datastrategiassaan, miten se aikoo hyödyntää dataa ja mitä dataa se tarvitsee. Mitä paremmin yritys tuntee kohderyhmänsä, sitä paremmin se menestyy. Kun tiedetään, mitä tukea kuluttaja tarvitsee ostoprosessissa tiedon hakemisen ja ostopäätöksen vaiheissa, on mahdollista vaikuttaa kauppohen saattamiseen päätökseen. B2B-sektorilla toimivat yritykset tarvitsevat tietoa asiakkaidensa asiakkaista eli kuluttajista. Tällöin tarvitaan dataa yrityksen ulkoisista lähteistä, jolloin data voi olla avointa tai maksullista. EU:n alueella toimitaessa yleinen tietosuoja-asetus (GDPR) asettaa reunaehdot kuluttajien tietojen keräämiselle, käsittelemiselle ja käyttämiselle.

Datan tyypit ja muodot

Dataa voidaan kerätä sekä sisäisistä että ulkoisista lähteistä. Yleensä sisäinen data on helpommin ja edullisemmin saavutettavaa, koska liiketoiminta omistaa ja hallinnoi sitä. Ulkoista dataa hallinnoi julkinen tai yksityinen organisaatio. Julkinen data on yleensä avointa, ja sitä voi ladata ilmaiseksi. Sen sijaan yksityinen data tulee ostaa suoraan yritykseltä tai kolmannen osapuolen datan toimittajalta. (Marr 2016, xii–xiii.)

Data voi olla tyypiltään strukturoitua, semistrukturoitua tai strukturoimatonta (taulukko 8). Strukturoitu data on järjestynyttä ja sijaitsee tietueen tai tiedoston sisällä kentässä esimerkiksi relaatiotietokantojen tai taulukkolaskennan datat. (Marr 2016, xii–xiii.) Strukturoitua dataa voi toisin sanoen tallentaa tietokantaan tai taulukkoon, ja tyypiltään se on määrällistä, kuten hinta tai kategorista, kuten sukupuoli (Taanila 2019b, 1). Sääntöjä noudattavana sitä on helppo syöttää, varastoida ja analysoida. Arviolta 80 % liiketoiminnan kannalta oleellisesta informaatiosta on kuitenkin alkuperältään strukturoimatonta tai semistrukturoitua dataa. Strukturoimaton data on ihmisen tuottamaa esimerkiksi tekstiä, valokuvia (Marr 2016, xii–xiii.), puhetta tai videoita (Taanila 2019a, 1). Sen ymmärtäminen on haastavaa tietokoneelle (Sharda 2018, 83). Semistrukturoitu data on strukturoidun ja

strukturoidun datan välimuoto, jolloin jotkin rakenteet ovat analysoitavissa (Marr 2016, xii–xiii).

Massadata eli big data ilmiönä viittaa datan määrään, vaihteluun, vauhtiin, todenmukaisuuteen ja arvoon. Volyymi kertoo datan suuresta määrästä ja vaihtelu datan erilaisista muodoista ja laadusta. Vauhti viittaa datan syntymisen ja liikkeen nopeuteen, ja totuudenmukaisuus datan synnyn monimutkaisuuteen, joka voi aiheuttaa epävarmuutta tuloksiin. Arvo tarkoittaa, että massadatan tulee tuottaa arvoa ja olla hyödynnettävää (Marr 2016, xii.) Esimerkiksi Google Trendistä kerätty data täyttää massadatan kriteerit.

Digitaalisen analytiikan tuottamat datat

Digitaalisen analytiikan tuottamat datat on kerätty verkosta, sosiaalisesta mediasta, hakukoneista tai muista digitaalisista lähteistä (Sponder & Khan 2018, xvi). Digitaalinen analytiikka tuottaa kolmenlaista dataa: ensimmäisen, toisen ja kolmannen osapuolen dataa (kuvio 8). Datat eroavat toisistaan sen mukaan, miten ne on tuotettu, hankittu ja prosessoitu (Sponder & Khan 2018, 71).

Ensimmäisen osapuolen data

Ensimmäisen osapuolen data on erittäin arvokasta, koska se kerätään suoraan omalta yleisöltä tai asiakkailta. Koska yritys omistaa datan itse, on sen hallinta ja tietosuojasta huolehtiminen mutkatonta. (Sponder & Khan 2018, 71.) Verkkoanalytiikalla kerätty data on esimerkki ensimmäisen osapuolen datasta. Se kerätään omilta verkkosivuilta asettamalla evästeitä vierailijoiden selaimiin. (Sponder & Khan 2018, 74.) Isobritannialaisen tutkimuksen mukaan lähes 80 % verkossa surffaavista kuluttajista on kyllästynyt mainosten tulvaan. Jo nykyään Firefox- ja Safari-selaimet estävät automaattisesti evästeiden käytön. Google aikoo lopettaa kolmansien osapuolien evästeet Chrome-selaimessa vuoteen 2022 mennessä. (Turunen 2020.)

Muita esimerkkejä ensimmäisen osapuolen datasta ovat sähköpostilistat, ostodatat yrityksen tietokannassa ja sosiaalisen median kanavien data (Sponder & Khan 2018, 71). Tarkkaa ja merkityksellistä ensimmäisen osapuolen dataa voi hyödyntää esimerkiksi tulevaisuuden ennustamiseen, asiakkaiden analysoimiseen, sisällön ja mainonnan personointiin ja mainonnan uudelleen kohdentamiseen (Relevant 29.11.2019). Heikkoutena on ulkopuolisen kilpailija- ja toimialadatan puute, mutta esimerkiksi Google Analytics tarjoaa anonyymia dataa omien sivujen toimialaa ja aluetta koskevaan benchmarkkaukseen (Sponder & Khan 2018, 74).



Kuvio 8. Ensimmäisen, toisen ja kolmannen osapuolen datalähteet (mukaillen Sponder & Khan 2018, 71, 72; Relevant 29.11.2019)

Toisen osapuolen data

Toisen osapuolen data kerätään myyntiin suoraan muilta organisaatioilta, yrityksiltä ja yksilöiltä. Esimerkiksi pankit keräävät tietoja luottokorttien käyttäjien toiminnasta ja jakavat niitä kumppaneille. Datan ostava yritys voi kontrolloida rajoitetusti datan merkityksellisyttä tai syvyyttä, koska kerääjä jakaa sen useiden yhteistyökumppaneiden kanssa. Toisaalta datan lähde on tiedossa, ja sen tarkkuus on luotettavaa. (Sponder & Khan 2018, 72; Relevant 29.11.2019). Toisen osapuolen datalla voi muun muassa laajentaa arvokasta omaa dataa, saavuttaa uusia asiakkaita ja rakentaa asiakkuuksia, ennustaa käyttäytymismalleja sekä auttaa yritystä kasvamaan (Relevant 29.11.2019).

Kolmannen osapuolen data

Kolmannen osapuolen dataa myyvät yritykset, jotka eivät ole alun perin itse keränneet dataa esimerkiksi markkinatutkimusyritykset. Ne keräävät suuria määriä dataa useista lähteistä, jalostavat sitä muun muassa tilastollisilla menetelmillä ja myyvät eteenpäin. (Sponder & Khan 2018, 72; Relevant 29.11.2019). Kolmannen osapuolen dataa käytetään kohdentamiseen. Datan ostaja ei tiedä datan lähdettä ja luotettavuutta, ja sama data on myös kilpailijoiden ulottuvilla. Kolmannen osapuolen dataa voi hyödyntää esimerkiksi rikastamaan omaa dataa sekä laajentamaan ja löytämään uusia yleisöjä. (Relevant 29.11.2019.)

Google tarjoaa toisen ja kolmannen osapuolen dataa DoubleClick Ad Networkista. Dataa tarjoavat myös esimerkiksi IBM Coremetrics ja Adobe Analytics, johon yritykset ovat rakentaneet rajapintoja Comscoren ja Nielsenin kolmannen osapuolen dataa varten. (Sponder & Khan 2018, 74.)

Ulkoisen datan keräämismenetelmät Internetistä

B2B-yrityksellä on monia keinoja lisätä kuluttajaymmärrystä ostopäätösprosessista yrityksen ulkoisen datan avulla. Ulkoisen datan hankkimiseen Internetistä on kolme menetelmää: paneeli- ja ISP-pohjainen mittaaminen sekä hakukonedatan kerääminen (Kaushik 2007, 44).

Paneelipohjainen mittaaminen

Paneelidataa kerää ja analysoi monikansallinen Comscore-yritys, joka tuottaa markkinadataa digitaalisten palveluiden yleisömääristä, käytöstä ja kuluttajakäyttäytymisestä. Data pääsee syvälle kuluttajien nettikäyttäytymisessä, mutta otoskoko voi olla suhteellisesti melko pieni (Kaushik 2007, 44–45.) Comscore kerää muun muassa tagittamalla census-dataa, paneelissa mukana olevien henkilöiden tietokoneen käyttöä mittaavan ohjelmiston avulla paneelidataa, mobiilikäytön demografiatietoja ja väestöestimaatit eri ikä- ja sukupuoliryhmissä päätelaitteiden käyttötutkimuksella (FIAM 2020a).

Paneelidata koostuu census-datasta ja paneelidatasta. Census-datalla saadaan realistinen kuva liikenteen määrästä, ja paneeli tuottaa tiedot profiileista. Yhdistämällä tiedot saadaan mallinnettua jokaiselle evästeelle demografiatieto. (FIAM 2020a.) Paneelin jäsenet ovat antaneet luvan nettisurffailun seurantaan vastiketta vastaan, joka voi olla vaikka serveripohjainen virussuoja (Kaushik 2007). Esimerkiksi ilmainen Avast-virustutka myi muun muassa Googlelle vuoden 2020 alkupuolelle asti selainlaajennuksilla keräämäänsä tietoa, jolla käyttäjien liikkumista verkossa saattoi seurata hyvinkin tarkasti. Yksittäisen käyttäjän tunnistaminen tietojen perusteella oli myös mahdollista päätelemällä. Tällä hetkellä Avast kerää tietoja virustutkan kautta. (Linnake 2020.)

Suomessa Media Metrics Finland Oy:n tuottama FIAM (The Finnish Internet Audience Measurement) mittaa suomalaisten medioiden sivustojen ja sovellusten liikenteestä noin 95 %. Se on online-yleisöjen virallinen mittaus, jossa raportoitavia tunnuslukuja on 20 muun muassa uniikit vierailijat, päivittäiset vierailijat keskimäärin, sivunäytöt ja keskimääräinen vierailun kesto minuutteja. Data muodostuu census- ja paneelidatasta. Census-data näyttää liikenteen määrän ja soveltuu kävijämäärän laskentaan. Paneelidata kerätään paneeleista. Paneelissa on vähintään 100 kohderyhmiin kuuluvaa henkilöä eri sukupuoli- ja ikäryhmistä (15-24, 25-34, 35-44, 45-54 ja 55-64). Tietoja on saatavana hyvin

monipuolisesti niin tunnusluvuista kuin demografisista tiedoista eri raportointiliittymistä (FIAM 2020b.) Esimerkiksi asiakassegmentti- ja ostodataan yhdistetty asiakaspaneelidata tuo lisätietoa asiakkaan mieltymyksistä ja käyttäytymisestä (Salo 2014, 38–42).

Nielsen tutkii kuluttajapaneeliin liittyneiden henkilöiden päivittäistavaraostoksia kuluttajien skannaamien kuittien perusteella ja tekee kyselyitä. Tutkimus kattaa kuluttajien kiinnostuksen kohteet, ostot, käyttäytymisen ja mieltymykset. Kuluttajapaneeli muodostuu erilaisista suomalaisista talouksista vastaajien perhemuodon ja iän mukaan. Panelistien skannamat datat kootaan yhteen anonymisti ja valmistellaan myytäväksi kaupan ja teollisuuden käyttöön. Kaikkiaan paneelissa on noin 5 000 henkilöä. API-rajapintoja on lukuisia. (Nielsen 2020.)

ISP-pohjainen mittaaminen

ISP (Internet Service Providers) -pohjainen mittaaminen käyttää anonymia dataa, joka on kaapattu erilaisilta verkkopalveluiden tarjoajilta. Otokoko on suuri ja tiedon tarjonta laavaa, mutta tieto ei mene syvälle nettikäyttäytymisessä. Hitwisein kaltaiset yritykset keräävät sopimukseen perustuen verkkopalveluiden tarjoajilta lokitietoja, jotka ne sitten analysivat. Data yhdistetään paneelitietoon demograafisten ja elämäntapatietojen tuottamiseksi. (Kaushik 2007, 46.)

Hakukoneet

Kuluttajan ostoprosessin tiedon etsiminen -vaiheesta on saatavana hakukoneista avointa dataa, jota voi kerätä myös useilla maksuttomilla hakukoneanalytiikan työkaluilla, kuten Google Trends tai Ubersuggest.

Hakukonedata muodostuu hakukoneiden massiivisista hakumääristä. Markkina- tai kilpailijaymmärryksen kartuttamiseen tarkoitetuilla työkaluilla voi seurata esimerkiksi minkä vain domain-osoitteen tai sovelluksen trafiikin määrän kehittymistä. Usein hakukoneilla on tietoa myös käyttäjistä. Esimerkiksi Google Trendsistä voi kerätä hakusanoihin perustuvaa markkinatietoa. (Kaushik 2007, 47–48.)

Muut ulkoisen datan lähteet

Kun halutaan lisätä ymmärrystä kuluttajan ostoprosessin ostopäätösvaiheesta, tarvitaan tietoa myynneistä tai ostoista. B2B-yritys ei voi hyödyntää yrityksen sisäistä transaktiodataa, joten tarvitaan muita tapoja hankkia tietoa. Päivittäistavara- ja pankit keräävät dataa myydyistä tuotteista ja ostavista asiakkaista, ja tilastokeskuksen keräämät kulutusdatat voivat olla käyttökelpoisia joissakin tapauksissa. Datojen hyödyntämisen esteeksi muodostuu kuitenkin yleensä saatavuus.

Avoim elintarvikkeiden kulutustieto

Tilastokeskuksen toteuttama kulutustutkimus kerää kotitalouksilta tietoa muun muassa mitä elintarvikkeita kuluttajat hankkivat ja kuinka paljon. Tutkimus perustuu otokseen, jossa tiedot kerätään haastattelemalla, päiväkirjoista, kauppaostosten kuiteista ja hallinnollisista rekisteriaineistoista. Elintarvikkeiden määrätiedot arvioidaan yksikköhintojen ja kuluttajan päiväkirjassa ilmoittamien tietojen perusteella. Tiedot tallennetaan keskiarvoina kg/hlö/vuosi. Tutkimuksessa tarkastellaan lisäksi erilaisten kotitaloustyyppien välisiä eroja esimerkiksi nuorten ja vanhempien tai naisten ja miesten välillä elintarvikkeiden hankinnassa. (Aalto 2018, 6, 8, 9, 10.) Taustatiedot ovat saatavana tarkimmalla luokituksella muun muassa viitehenkilön iän ja sosioekonomisen aseman sekä kotitalouden tyypin, asuinalueen ja tulojen perusteella (Tilastokeskus 2018b). Kulutustutkimus on ainoa tietolähde Suomessa, jonka sopii väestöryhmien ja alueiden välisten kulutuserojen tarkasteluun. Lähes puolet kulutustutkimuksen tiedoista on päivittäistavaraostoja, joiden tiedot saadaan kotitalouksien kahden viikon aikana keräämistä kuiteista. Tilastokeskuksessa kuitit skannataan ja digitoidaan. (Hatakka 2015.) Kehitteillä on tietojen keräämiseen mobiilipohjainen sovellus, jolla kuluttaja skannaa kuitit ja automaattinen tekstintunnistus lukee tuotteet kuiltalta ja luokittelee automaattisesti (Kajantie 18.11.2019). Tilastokeskuksen avoimet tietokanta-aineistot voi noutaa PxWeb API -rajapinnasta (Tilastokeskus 2020b).

K-ryhmän asiakas- ja asiointitieto

K-ryhmän K-Ruoka ja verkkokaupparekisteri kerää henkilö- ja asiointitietoja. Henkilötiedoista kerätään mm. nimi, osoite, puhelinnumero ja sähköpostiosoite ja verkkokauppa-asiointista verkkoselailu- ja ostotiedot, tilaukset, toimitukset, palautukset sekä suosikkikaupat ja -tuotteet. Henkilötietoja voidaan luovuttaa yhteistyöyrityksille palveluiden toteuttamiseksi. (Kesko Oyj 2020a.) Tietoja ei myydä ulkopuolisille yrityksille (Lindberg 22.1.2020). Jos tietoja myöhemmin halutaan käyttää uusiin tarkoituksiin, asiakasta informoidaan käsittelyperusteesta ja pyydetään tarvittaessa suostumus. Muun muassa poliisi-, tulli-, rajavartiolaitos- ja veroviranomaisilla on lakiperusteinen tiedonsaantioikeus. (Kesko Oyj 2020a.)

K-plussan asiakasrekisteri kerää seuraavat henkilötiedot: nimi, osoite, puhelinnumero, sähköpostiosoite, syntymäaika tai henkilötunnus, sukupuoli ja kieli. Asiakkuustiedoista kerätään muun muassa asiakasohjelmiin kuulumisen, käyttöönnotetut sovellukset ja palvelu ja Plussa-kortin tiedot. Ostotiedoista tallennetaan asiakkaan sallimalla tasolla loppusumma, tuoteryhmä, tuote sekä ostokäyttäytymislukko. Rekisteristä voidaan luovuttaa tietoja Plussa-ohjelmaan kuuluville yrityksille markkinointitarkoituksiin ja asiakasrekisterin rekisteritietojen päivitykseen. Jos MobilePay-palvelua käyttöönottaessa asiakas sallii tietojen luovuttamisen, voidaan palveluntarjoajalle luovuttaa tietoja. Jos tietoja myöhemmin

halutaan käyttää uusiin tarkoituksiin, asiakasta informoidaan käsittelyperusteesta ja pyydetään tarvittaessa suostumus. Muun muassa poliisi-, tulli-, rajavartio- ja veroviranomaisilla on lakiperusteinen tiedonsaantioikeus. Sähköinen markkinointi, tutkimus, asiakasviesintä, asiakasrekisteritietojen luovutus Plussa-ohjelmaan kuuluvien yritysten asiakasrekisteritietojen päivitykseen, profilointi ja ostojen rekisteröiminen on kiellettävä erikseen. (Kesko Oyj 2020b.)

S-ryhmän asiakas- ja asiointitieto

S-ryhmän asiakasomistaja- ja asiakasrekisteriin kerätään muun muassa seuraavat henkilötiedot: nimi, osoite, puhelinnumero, sähköpostiosoite, syntymäaika, henkilötunnus, sukupuoli, kieli ja turvakielto. Lisäksi talletetaan tiedot muun muassa osuuskaupan jäsenyydestä, rekisteröitymiset bonusyhteistyöyrityksiin, S-etukortin tiedot, käyttöönotetut palvelut ja käyttö sekä verkkosivujen ja mobiilipalveluiden käyttäjätiedot. Ostotiedoista tallennetaan muun muassa S-etukortin numero tai jäsennumero, bonus sekä kortin käyttötapa, ostopäivä, kellonaika, ostopaikka, tiedot ostoista kuitenkin loppusumma-, tuoteryhmä- ja/tai tuotetasolla. Verkkokauppa-asioinnissa kerätään lisäksi tiedot tallennetusta ostoskorista, tilaus- ja toimitustiedot sekä asiakassegmentti. Esimerkiksi tuotetasoisen ostotiedon, puhelinmarkkinoinnin, tutkimuskyselyjen ja sähköpostitiedotteiden esto on tehtävä erikseen. Tietoja voidaan käyttää S-ryhmän yhteisöjen ja asiakasomistajajärjestelmään kuuluvien yhteistyökumppaneiden markkinointiin. (SOK 2020.) S-ryhmän asiakastietoja ei kuitenkaan luovuteta yhteistyökumppaneille, vaan esimerkiksi mainonta lähetetään suoraan painotalosta (Raivio 2018). Viranomaisille tietoja luovutetaan asiakasrekisteristä lainsäädännön vaatimusten mukaan (SOK 2020).

Lidlin asiakas- ja asiointitieto

Lidl on ottamassa mobiilisovelluksena toimivan Lidl Plus -etuohjelman käyttöön vuoden 2020 aikana. Ohjelman käyttöönotto vaatii Lidl-tilin rekisteröimisen, jolloin kerätään muun muassa asiakkaan nimi, syntymäaika, puhelinnumero, sähköposti ja myymälä, jonka tarjoustarjontaa asiakas haluaa seurata. Asiakkaan ostokuitit tallentuvat palveluun. (Rissanen 2020.)

Päivittäistavarakaupan kanta-asiakasdatan hyödyntäminen tilastoinnissa

Suomessa tilastokeskus tutkii mahdollisuutta hyödyntää yksityisen sektorin aineistoja kotitalouksien muun muassa elintarvikkeiden kulutuksen tilastoinnissa, jolloin kanta-asiakasrekisteritiedoilla voitaisiin korvata kuittien kerääminen kotitalouksissa. Suomen päivittäistavarakauppa on keskittynyt: kaksi suurinta kauppaketjua, S- ja K-Ryhmä, kattavat noin 80 % markkinaosuudesta. Ostotapahtumat tulee rekisteröidä kulutustutkimusta varten tuo-

tetasolla COICOP-luokituksen mukaan. Ostotiedoissa tulisi olla myös elintarvikkeiden tuotekohtainen määrätieto. Tietojen saaminen kaupparyhmiltä edellyttää kotitalouksien lupaa yksilötasolla. (Hatakka 2015.) Asiakaskorttiaineistot kattavat ainoastaan osan kotitalouksien päivittäistavaraostoista. Vuonna 2016 kulutustutkimuksen kuiteista kaksi kolmesta oli lähtöisin S- ja K-ryhmän liikkeistä. Arvion mukaan suurten ketjujen myynnistä asiakaskorttiosiot ovat noin 75 %, jolloin vain noin puolet ostoista on mahdollista jäljittää. Kanta-asiakaskorttien ostodata on vinoutunut, koska nuoremmat kuluttajat käyttävät vähemmän kanta-asiakaskortteja. Osittainen aineisto tulee täydentää, jotta se olisi käyttökelpoista. (Kajantie 18.11.2019.) Lidl on lanseeraamassa mobiilisti toimivaa asiakkaiden etuohjelmaa vuonna 2020. Lidlin markkinaosuus päivittäistavara kaupasta on noin 10 %, joten ostojen jäljitettävyyسمahdollisuudet paranevat tulevaisuudessa. (Järvinen & Brännare 2020.)

Ostaja- ja tuotedatan yhdistäminen kaupan ja pankin aineistoista

Tilastokeskus kartoittaa mahdollisuutta kerätä tietoa yksityisen sektorin aineistoista yhdistämällä pankkien transaktiodata ja kauppojen maksupääteaineistot, joista kumpikaan ei sovellu yksinään tietolähteeksi. Tämä johtuu siitä, että pankkien transaktiodatassa ei ole riittävän tarkasti tuotetasolle menevää tietoa ja kauppojen maksupääteaineisto ei puolestaan sisällä tietoja ostajasta. Norjan tilastovirasto on testannut tiedonkeruutapaa yhden kauppaketjun maksupääteaineistolla yhden päivän ajan. Yhdistelyssä käytettiin aika- ja yritystietoa sekä oston loppusummaa. Tietojen yhdistämistä edesauttoi Norjan keskittynyt maksukorttijärjestelmä, jossa maksukorttiliikennettä hoitaa yksi toimija. (Kajantie 18.11.2019.) Massadatan käyttö parantaa tiedon oikeellisuutta. Dataa on mahdollista kerätä pikkuhiljaa, jolloin voidaan välttää kausivaihtelua. Keskeiset laatuongelmat tämän tyyppisessä datassa ovat valikoivuus ja potentiaalinen kattavuusvirhe. (Haraldsen ym. 2014.) Haasteena Suomessa on maksukorttiliikenteen toimijoiden lukumäärä, massiivisten aineistojen kerääminen ja käsittely sekä tiedonkeruuta koskevien asetusten muutokset (Kajantie 18.11.2019). Bruun ja Mägi (2020) toivovat, että tilastoinnin pohjaksi saataisiin käyttöön yritysten tietoaaineistoja, kuten päivittäistavara kaupan kassatapahtumat, kanta-asiakasdatat ja mobiilioperaattoreiden paikannusdatat. Datan saatavuus hyödyttäisi epäilemättä myös monien yritysten liiketoiminnan kehitystä.

Avoimet myyntidatat

Systemaattisesti kerättävää ja julkaistavaa avointa dataa elintarvikkeiden myynnistä on saatavana ainoastaan Alkon alkoholijuomien myyntiraporteista, jotka julkaistaan kuukausitasolla Suomesta ja eri maakunnista (Alko 2020). Myynnit ilmaistaan litroina. Tiedot ovat pdf-muodossa.

Datan analysointi

Analytiikalla tarkoitetaan datan laajaa käyttöä, tilastollista ja kvantitatiivista analysointia, selittäviä ja ennustavia malleja sekä tietoon perustuvaa päätöksentekoa ja toimintaa. Analytiikka voi olla syötteenä inhimillisille tai täysin automatisoidulle päätöksille. Ennustava analytiikka hyödyntää kvantitatiivisia menetelmiä, kuten ekonometristä ennustamista ja malleja, jotka käyttävät historiatietoja tulevaisuuden ennustamisessa (Davenport & Harris 2017, 25). Analytiikkaohjelmisto on yksikertaisimmillaan taulukkolaskentaohjelma, kuten Excel, jossa on tilasto- ja optimointityökaluja. Perinteisiä tilasto-ohjelmia ovat esimerkiksi Minitab ja Stata. Monimutkaisia datan visualisointi- ja kuvailevan analytiikan ohjelmistoja ovat Qlik, Tableau, MicroStrategy, Oracle Hyperion ja IBM Cognos. Laajoja ja monipuolisia ennustavan- ja ohjaavan analytiikan ohjelmistoja tarjoavat esimerkiksi SAS ja IBM sekä perinteiseen tilastoanalyysiin että massiivisten strukturoimattomien datojen analysointiin. (Davenport & Harris 2017, 26–27.)

Data-analytiikka

Data-analytiikka on prosessi, jossa tietojoukkojen sisältämää informaatiota tarkastellaan johtopäätösten tekemiseksi yleensä tietojärjestelmien ja ohjelmien avulla. Data-analytiikan teknologioita ja tekniikoita käytetään laajasti kaupallisilla aloilla tietoon perustuvassa päätöksenteossa. Data-analytiikka on laaja käsite. Siinä missä liiketoiminta-analytiikka suuntautuu analysoimaan dataa liiketoiminnan tarpeisiin, data-analytiikka kohdistuu laajempaan toimintaympäristöön. Analysoitava data voi sisältää historia- ja uutta tietoa, jotka prosessoidaan reaaliaikaisen analytiikan tuottamiseen. Data voidaan koota sisäisistä järjestelmistä ja ulkoisista datalähteistä. (Rouse 2017.)

Massadata-analytiikka käyttää tiedonlouhinnan, prediktivisen analytiikan ja koneoppimisen työkaluja isoihin tietojoukkoihin, jotka usein sisältävät sekä strukturoimatonta että semistrukturoitua dataa (Rouse 2017). Usein massadata halutaan jalostaa automatisoiduksi päätöksenteoksi, jolloin nykytilanne tiedostetaan reaaliaikaisesti ja pystytään ennustamaan tulevaisuutta. Pilvipalvelut mahdollistavat teknisesti tehokkaan tavan tallentaa, jakaa ja jalostaa käsittämättömän suuria datamääriä. (Salo 2014, 26–32.) Massadata ja sen analysointi ovat osa kokonaisvaltaista liiketoiminnan kehittämistä, joka tukee päätöksentekoa ja synnyttää lisäarvoa muun muassa tuomalla kilpailuetua ja lisäämällä asiakasymmärrystä. Massadatan isot volyymit vaativat sopivia työkaluja datan analysointiin. Esimerkiksi Hadoop ja Apache Spark selviytyvät massiivisista datamääristä. (Mueller & Massaron 2018, s. 123.)

Digitaalinen analytiikka

Kaushikin mukaan digitaalinen analytiikka on kvalitatiivisen ja kvantitatiivisen liiketoiminta- ja kilpailijadatan analysointia, jolla ohjataan asiakkaiden verkkokokemuksen jatkuvaa parantamista kohti toivottua lopputulosta. Digitaalinen analytiikka perustuu ajatukselle, että asiakas on markkinointikonseptin keskiössä ja voi aloittaa ostopolun mistä tahansa ostopäätöspolun kohdasta. Markkinoijan tehtävä on ennakoida, millaisia viestejä asiakas haluaa kuulla. (Sticky Media 2020.) Google-hakukone ja -verkkopalvelut mullistivat tavan saavuttaa ja käyttää dataa helpolla tavalla (Sponder & Khan 2018, 1). Digitaalinen analytiikka voidaan jakaa ulkoiseen ja sisäiseen analytiikkaan (taulukko 8). Ulkoisella analytiikalla analysoidaan kilpailijoita tai markkinoita esimerkiksi SimilarWeb- ja Google Trends -työkalujen avulla. Sisäisellä analytiikalla analysoidaan puolestaan oman verkkosivuston ja sosiaalisen median profiilien dataa liiketoiminnallisesta lähtökohdasta esimerkiksi CRM:llä tai Google Analyticsillä (Salminen 2.4.2015.)

Hakukoneanalytiikka

Hakukoneanalytiikalla tarkoitetaan hakukoneoptimointia ja hakutrendien analysointia. Hakukoneoptimoinnin tarkoitus on optimoida verkkosivut sellaisiksi, että suurin osa vierailijoista löytää ne hakukoneiden avulla, tuottaa sivuille trafiikkia ja nostaa sivut hakutuloksissa mahdollisimman korkealle. Hakukoneiden tuottamat tulokset ovat joko orgaanisia (ei-maksullisia) tai maksettuja. Hakukoneet järjestävät orgaaniset hakutulokset algoritmin yli 200 tekijän perustella. (Sponder & Khan 2018, 19, 24, 34.) Google kattoi maailmanlaajuisesti lähes 93 % vuonna 2019 tehdyistä hauista (Oberlo 2020).

Google.com on yleinen hakukone (Crawler-Based Search Engine), joka kerää ja varastoi informaation verkkosivuilta indeksointia varten. Indeksointi auttaa luokittelemaan sivuston hakua varten, jolloin data on saatavana nopeasti tietokannasta. Kun käyttäjä näppäilee hakusanat, hakukone tiedustelelee indeksiä ja tarjoaa listan osuvimmista verkkosivuista vertaamalla indeksoituihin hakusanoihin. Lista vaikuttaa myös hakualgoritmien määrittämä sisällön laatu (On Page), paluulinkit muilta verkkosivuilta, hakukoneista, sosiaalisesta mediasta ja sähköpostista (Off page) sekä metadataan ja käytettävyyteen liittyvät asiat, kuten esimerkiksi otsikko, kuvaus, avainsanat, saavutettavuus mobiilisti ja sivujen latausnopeus. (Sponder & Khan 2018, 20–21.)

Taulukko 8. Esimerkkejä digitaalisen analytiikan kohteista, työkaluista sekä datatyypeistä ja muodoista (mukaillen Crestodina 17.7.2017; Herrmann 19.10.2018; Marr 2016, xiii; Salminen 2.4.2015)

Digitaalinen analytiikka	
SISÄINEN ANALYTIKKA Verkko- ja sosiaalisen median analytiikka	ULKONEN ANALYTIKKA Kilpailija- ja markkina-analytiikka
Kohteet	
Omat verkkosivut Omat sosiaalisen median profiilit	Kilpailijat Markkinat
Työkalut	
Google Analytics, Matomo, Mixpanel, SimilarWeb, CrazyEgg, Hotjar, Lucky Orange, Facebook Analytics, Instagram Insights, Twitter Analytics, quintly, CRM	Google Trends, Ubersuggest, SimilarWeb, Google Alerts, Ahrefs, Alexa, Searchmetrics, SpyFu, MOZ, SEMrush, Meltwater
Datatyypit ja muoto	
Asiakasdata Yritysdata	Kuluttajadata Markkinadata Kilpailijadata
Strukturoitu	
Myyntidata Talousdata esim. transaktiohistoria Asiakasdata Henkilöstökyselyt HR-tietueet	Markkinatutkimus ja trendidata Väestönlaskentainformaatio GPS-sijaintidata Sensoridata Säädata
Semistrukturoitu	
Tägätyt tai kategorisoidut valokuvat, graafit ja videot Sähköposti	Tägätyt valokuvat ja videot Järjestetty grafiikka Kategorisoitu teksti
Strukturoimaton	
Verkkosivut Tekstitiedostot Valokuvat Ääni Sosiaalinen media Asiakaspalaute	Verkkosivut Tekstitiedostot Valokuvat Ääni Sosiaalisen median profiilidata

Hakutrendien analysoinnilla hankitaan ymmärrystä, millä avainsanoilla ihmiset tekevät hakuja tietynä ajanjaksona (Sponder & Khan 2018, 24–27). Toimivat avainsanat (hakusanat, hakutermit) liittyvät oleellisesti tuotteeseen, tähtäävät liiketoiminnalliseen tavoitteeseen ja niillä on suuri hakuvolyymi halutun yleisön joukossa. Avainsanoja valitessa on tärkeää kiinnittää huomiota sanojen ja myytävän palvelun tai tuotteen väliseen yhteyteen. Niiden tulisi vastata mahdollisimman hyvin potentiaalisten asiakkaiden käyttämiä hakusanoja. Sanoina kannattaa käyttää niin sanottuja long tail -hakusanoja, jolloin hakusana muodostuu useammasta, yleensä kahdesta tai kolmesta erillisestä sanasta tai lauseesta.

Oikein valittujen hakusanojen avulla yrityksellä on suurempi mahdollisuus päästä Googlen hakutulosten kärkeen. (Rouhiainen 2019.) Google Trends ja Ubersuggest ovat esimerkkejä hakukoneanalytiikan työkalusta.

Verkkoanalytiikka

Verkkoanalytiikka on verkkosivujen trafiikin ja sivuston konversion mittaamista ja analysointia. Sen tarkoitus on lisätä ymmärrystä vierailijan käyttäytymisestä sivuston optimoimiseksi konversioprosessin, navigoinnin, kampanjoiden ja estetiikan suhteen. (Sponder & Khan 2018, 83.) Verkkodatan avulla pyritään parantamaan asiakaskokemusta ja lisäämään liiketoiminnallista hyötyä kuten myyntiä (Sponder & Khan 2018, 13). Google Analytics -työkalulla voi seurata yrityksen omien verkkosivujen trafiikkia ja verkkokaupan transaktiotapahtumia. Transaktiodata tarkoittaa dataa, joka syntyy yrityksen liiketoimintaprosesseissa esimerkiksi ostot ja tilaukset, ja niistä kertyviä datamääriä (Finanssiala 2019). Asiakkaiden transaktiotiedot voi hakea helposti ja automaattisesti tuotteille luotujen uniikkien koodien (SKU) avulla Google Analyticsilla reaaliaikaisesti. Google Analyticsista automaattisesti poimittua myyntidataa voi hyödyntää esimerkiksi ennustemalleissa. Trafiikkitiedoista näkee vierailijat, mistä vierailijat tulevat sivuille, mitä sivuja selaavat ja konvertoinnit. (Boone, Ganeshan & Hicks 2015.)

EU:n yleinen tietosuoja-asetus (GDPR)

EU:n alueella henkilötietojen käsittelyä säätelee yleinen tietosuoja-asetus (GDPR). Asetuksen tarkoitus on suojata kansalaisten henkilötietoja ja tarjota keinoja hallita omien tietojen käsittelyä. Kuluttajalla on asetuksen perusteella muun muassa oikeus tietää, mitä henkilötietoja organisaatio kerää, miten niitä käsitellään ja mihin tarkoitukseen käytetään sekä vastustaa käsittelyä, pyytää rajoittamaan tai poistamaan henkilötietoja. Yrityksellä on oikeus käsitellä henkilötietoja laissa määriteltyjen perusteiden mukaan, jolloin peruste pohjautuu suostumukseen, sopimukseen, lakisääteiseen velvoitteeseen, elintärkeiden etujen suojaamiseen, yleiseen etuun ja julkiseen valtaan. Lisäksi se voi olla rekisterinpitäjän ja kolmannen osapuolen oikeutettu etu, jolloin taustalla on merkityksellinen suhde esimerkiksi asiakkuus. Henkilötietojen käsittely voi olla oikeutettua esimerkiksi suoramarkkinoinnissa tai tilastoinnissa. (Tietosuojavaltuutetun toimisto 2020.)

6.2 Päivittäiselintarvikkeiden hakuajankohdat

6.2.1 Meijerituotteet/voi

Trendi ja kausiluonteisuus

Voin hakeminen Google Trends -indeksien perusteella oli viiden vuoden ajanjaksolla trendinä kasvusuuntainen. Sen sijaan viiden erillisen vuoden hauista nähdään, että voi ei ole tuotteena kausiluonteinen, koska sitä haetaan ympäri vuoden.

Kuukausi- ja viikkokohtaiset erot hauissa

Voita haettiin viiden vuoden aikana (2015–2019) keskimäärin eniten joului- ja kesäkuussa, mutta hakuja tehtiin melko tasaisesti ympäri vuotta (indeksit vähintään noin 40). Voin hakuindeksien keskiarvoissa ei ollut kuitenkaan merkitseviä eroja eri kuukausien välillä 5 %:n riskillä. Voita haettiin keskimäärin eniten viikkoina 50 (joulukuu), 41 (lokakuu) ja 24 (kesäkuu). Voin hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitseviä eroja esimerkiksi viikkojen 50 ja 2, 3, 5, 10, 31, 42 tai 47 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Kolmen kuukauden ajanjaksolla voita haettiin eniten lauantaina ja perjantaina (indeksi > 50). Hakuintensiteetti nousi torstaista alkaen viikonloppua kohti ja taittui lauantain jälkeen. Voin hakuindeksien keskiarvoissa oli maanantain ja lauantain välillä tilastollisesti merkitsevä ero 5 %:n riskitasolla.

Voin hakemisen viikon trendissä näkyi eroja viikonpäivien ja vuorokaudenaikojen välillä. Eniten voita haettiin seitsemän vuorokauden keskiarvojen perusteella iltapäivällä klo 17 (indeksi > 60) ja toiseksi eniten yöllä klo 3 (indeksi > 60). Kaikkiaan voita haettiin nousujohteisesti aamuneljästä illansuun hakuhiippuun klo 17 asti, jonka jälkeen haut kääntyivät laskuun noustakseen taas yöllä klo 1 alkaen. Voin hakuindeksien keskiarvojen välillä oli tilastollisesti merkitseviä eroja eri kellonaikoina. Esimerkiksi klo 1 ja klo 3, 16, 17 tai 19 sekä klo 3 ja klo 4, 6–8, 10 tai 14 eroavat tilastollisesti merkitsevästi 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Karttojen mukaan voin hakemisen volyymi on kasvanut viidessä vuodessa Suomessa. Väripalkeista nähdään, että esimerkiksi Lapissa voin hakemisen intensiteetti on vaihdellut eri vuosien välillä muita maakuntia enemmän, kun puolestaan Uusimaalla ja Päijät-Hämeessä vuosien välinen vaihtelu on ollut pientä.

Voin hakeminen ikäluokittain

Voita hakivat eniten 35–44-vuotiaat lähes 38 % osuudella. Hakuja ei ollut ollenkaan tai niiden määrä oli hyvin pieni alle 18- ja yli 65-vuotiaiden sekä 55–64-vuotiaiden ikäryhmissä.

Voin hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 12.

6.2.2 Meijerituotteet/jäätelö

Trendi- ja kausiluonteisuus

Jäätelön hakeminen Google Trends -indeksien perusteella oli viiden vuoden ajanjaksolla trendinä kasvusuuntainen. Vuosien 2015–2019 viiden erillisen vuoden hauista nähdään, että jäätelön haut painottuvat kevät–kesään, mutta sitä haetaan kuitenkin ympäri vuoden.

Kuukausi- ja viikkokohtaiset erot hauissa

Jäätelöä haettiin keskimäärin eniten touko-, kesä- ja heinäkuussa. Jäätelön hakuindeksien keskiarvoissa eri kuukausina oli tilastollisesti merkitseviä eroja esimerkiksi touko-, kesä- ja heinäkuun ja tammi-, helmi- ja maaliskuun tai elokuun ja joulukuun välisenä aikana 5 % riskillä. Viikot 20–23 (touko–kesäkuun vaihde) ja 26–29 (kesäkuun loppupuolelta heinäkuun loppupuolelle) olivat keskimäärin vilkkainta jäätelön hakemisen aikaa (indeksi > tai yhtä suuri kuin 60). Jäätelön hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitseviä eroja esimerkiksi viikkojen 20–23 ja 1–2, 7, 36, 39, 43 tai 45–52 aikana 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Kolmen kuukauden ajanjaksolla jäätelöä oli haettu eniten viikonloppuna eli perjantaina, lauantaina ja sunnuntaina (indeksi 55–lähes 60), mutta jäätelön hakuindeksien keskiarvoissa ei ollut eri viikonpäivien välillä tilastollisesti merkitsevää eroa 5 %:n riskitasolla.

Jäätelön hakemisen trendi vaihteli viikon ajanjaksolla viikonpäivien ja vuorokaudenaikojen mukaan. Eniten jäätelöä haettiin seitsemän vuorokauden keskiarvojen perusteella iltapäivällä klo 16–19 (indeksi > 60). Kaikkiaan jäätelöä haettiin nousujohteisesti aamukuudesta iltapäivän hakuhiippuun klo 16 asti. Jäätelön hakuindeksien keskiarvojen välillä oli tilastollisesti merkitseviä eroja eri kellonaikoina. Esimerkiksi klo 16–18 ja 1–11 tai 22 eroavat tilastollisesti merkitsevästi 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Jäätelön hakemisen intensiteetti on kasvanut Suomessa viiden vuoden aikana. Väripalkeista nähdään, että jäätelön hakeminen on kasvanut monissa maakunnissa esimerkiksi Kanta-Hämeessä huomattavasti kahden viimeisen vuoden aikana.

Jäätelön hakeminen ikäluokittain

Jäätelöä hakivat eniten 18–24-vuotiaat (40 %) ja hieman vähemmän 25–34- ja 35–44-vuotiaiden ikäluokat 30 % osuuksilla. Muut ikäluokat eivät hakeneet jäätelöä ollenkaan tai hakivat hyvin vähän.

Jäätelön hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 13.

6.2.3 Meijerituotteet/juusto

Trendi ja kausiluonteisuus

Juuston hakeminen Google Trends -indeksien perusteella oli viiden vuoden ajanjaksolla trendinä kasvusuuntainen. Viiden erillisen vuoden hauista nähdään, että vaikka juuston hakeminen painottuu joulun seutuun, sitä haetaan kuitenkin tasaisesti ympäri vuoden.

Kuukausi- ja viikkokohtaiset erot hauissa

Juustoa haetaan ylivoimaisesti eniten joulukuussa, mutta muutoin tasaisesti ympäri vuotta. Voin hakuindeksien keskiarvoissa oli merkitsevä ero joulukuun ja kaikkien muiden kuukausien välillä 5 %:n riskillä. Juustoa haettiin keskimäärin eniten viiden vuoden aikana (2015–2019) joulukuun viikkoina 50 (indeksi > 80) ja 51 (indeksi lähes 80) ja marraskuussa viikkona 48 (indeksi noin 60). Juuston hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitseviä eroja esimerkiksi viikkojen 50 ja 51 sekä kaikkien muiden viikkojen paitsi 18, 28, 45, 48 ja 52 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Kolmen kuukauden ajanjaksolla juustoa oli haettu eniten lauantaina ja sunnuntaina. Juuston hakuindeksien keskiarvoissa oli keskiviikon ja lauantain välillä tilastollisesti merkitsevä ero 5 %:n riskitasolla.

Juuston hakemisen seitsemän vuorokauden trendissä on eroa viikonpäivien ja vuorokaudenaikojen välillä. Eniten juustoa haettiin seitsemän vuorokauden keskiarvojen perusteella iltapäivällä klo 16–18 (indeksi > 60). Kaikkiaan juustoa haettiin nousujohteisesti aamukuudesta illansuun hakuhiippuun klo 17 asti. Juuston hakuindeksien keskiarvojen välillä oli

tilastollisesti merkitseviä eroja eri kellonaikoina. Esimerkiksi klo 16–18 ja klo 6–9 eroavat tilastollisesti merkitsevästi 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Juuston hakuintensiteetti kasvoi useimmissa maakunnissa viimeisen vuoden aikana. Yleisesti ottaen hakuintensiteetti vaihteli maakunnasta ja vuodesta toiseen.

Juuston hakeminen ikäluokittain

Juustoa hakivat eniten 35–44-vuotiaat (41 %) ja 18–24-vuotiaat (39 %). Hakuja eivät tehneet ollenkaan tai niitä oli hyvin vähän alle 18- tai yli 55-vuotiaat.

Juuston hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 14.

6.2.4 Valmisruoat/maksalaatikko

Trendi ja kausiluonteisuus

Maksalaatikon trendi on lievästi laskeva Google Trendsin viiden vuoden hakuindeksien mukaan. Viiden eri vuoden (2015–2019) indeksien perusteella haut näyttävät jakautuvat melko tasaisesti ympäri vuotta. Indekseissä näkyy voimakkaita hakupiikkejä, jotka osuvat eri vuosina eri aikoihin.

Kuukausi- ja viikkokohtaiset erot hauissa

Maksalaatikkovalmisruokaa haetaan tasaisesti kaikkina kuukausina, paitsi heinäkuussa, jolloin hakuja on hyvin vähän (indeksi noin 0). Ruoan hakuindeksien keskiarvossa oli merkitsevä ero heinäkuun ja joulukuun välillä 5 %:n riskillä. Maksalaatikkovalmisruokaa haettiin keskimäärin eniten viiden vuoden aikana (2015–2019) joulukuussa viikolla 50 (indeksi > 60). Ruoan hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitseviä eroja esimerkiksi viikkojen 50 ja 3, 8, 9, 11, 14, 17, 22, 23, 29, 32, 33, 39, 45 tai 47 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Kolmen kuukauden ajanjaksolla maksalaatikkovalmisruokaa oli haettu eniten maanantaina (indeksi noin 20), mutta ei juuri ollenkaan keskiviikkona (indeksi noin 0). Ruoan hakuindeksien keskiarvoissa ei ollut kuitenkaan tilastollisesti merkitsevää eroa eri viikonpäivien välillä 5 %:n riskitasolla.

Maksalaatikon seitsemän päivän hakutrendissä oli kaksi selkeästi erottuvaa huippua, kun hakeminen muutoin oli melko vähäistä. Maksalaatikkovalmisruokaa haettiin seitsemän

vuorokauden keskiarvojen perusteella noususuuntaisesti, mutta polveilevasti aamuseitsemästä iltapäivän hakuhiippuun klo 16 (indeksi noin 40) asti. Valmisruokaa ei ajateltu eikä haettu yöllä klo 2–4 (indeksi 0). Maksalaatikon hakuindeksien keskiarvojen välillä oli tilastollisesti merkitsevä ero klo 16 ja klo 21 välillä 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Eri vuosien välisiä eroja maakuntien maksalaatikkovalmisruoan hakuintensiteetissä ei voitu verrata, koska palvelu ilmoitti: ”Tuloksia ei voi näyttää, koska riittävästi kyselyä vastaavia tietoja ei löytynyt.”

Maksalaatikon hakeminen ikäluokittain

Maksalaatikkoa hakivat eniten 35–44-vuotiaat (49 %). Alle 18- ja yli 65-vuotiaat eivät hakenneet ruokaa tai hakuja oli hyvin vähän.

Maksalaatikon hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 15.

6.2.5 Valmisruoat/kinkkukiusaus

Kinkkukiusaus-termi valmistajien nimeen yhdistettynä viimeisen 5 vuoden ja vuosina 2015–2019, antoi indeksin arvoksi 0 kaikkien viiden hakuvuoden kohdalla. Tämä tarkoittaa, että riittävästi hakuja ei ollut kyseisillä hakutermeillä. Viimeisen viiden vuoden indeksejä ja kartta- ja alialuetietoja ei myöskään ollut saatavissa: ”Tuloksia ei voi näyttää, koska riittävästi kyselyä vastaavia tietoja ei löytynyt.” Google Trendsin [kinkkukiusaus]-hakutermillä aiheeseen liittyvät kyselyt käsittelivät usein valmistusohjeita esimerkiksi ”kinkkukiusaus joulukinkusta” tai ”kinkkukiusaus perunasta”.

6.2.6 Valmisruoat/kasvissosekeitto

Hakutermi [kasvissosekeitto Rainbow] jouduttiin tiputtamaan pois hakutermistä kokeilujen jälkeen, koska palvelu ilmoitti: ”Sivua näytettäessä tapahtui virhe.” Tästä huolimatta tuloksia ei saatu: ”Tuloksia ei voi näyttää, koska riittävästi kyselyjä vastaavia tietoja ei löytynyt.” Google Trends -palvelussa [kasvissosekeitto]-hakutermillä aiheeseen liittyviä kyselyitä olivat esimerkiksi ”paras kasvissosekeitto” tai ”juuressosekeitto”, jotka viittaavat ruokien koti-valmistukseen.

6.2.7 Kasviproteiinit/Nyhtökaura

Trendi ja kausiluonteisuus

Nyhtökauran hakeminen viiden vuoden ajanjaksolla Google Trends -indekseillä näyttää trendinä hieman laskevalta. Viiden erillisen vuoden hauista nähdään, että Nyhtökauraa haetaan melko tasaisesti ympäri vuoden. Hakupiikit eivät eri vuosina osu selkeästi samalle ajanjaksolle, joten ne eivät näytä liittyvän kausivaihteluun. Nyhtökaura tuotiin markkinoille toukokuussa 2016.

Kuukausi- ja viikkokohtaiset erot hauissa

Nyhtökauraa haettiin viiden vuoden aikana (2015–2019) keskimäärin melko tasaisesti eri kuukausina ympäri vuotta, mutta suurimmat mediaanit osuivat helmi–maaliskuulle ja syys–lokakuulle. Nyhtökauran hakuindeksien keskiarvossa oli merkitsevä ero lokakuun ja touko- tai kesäkuun välillä 5 %:n riskillä. Nyhtökauraa haettiin keskimäärin eniten maaliskuussa viikolla 11 (indeksi lähes 40). Tuotteen hakuindeksien keskiarvoissa eri viikkoina ei ollut tilastollisesti merkitseviä eroja 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Kolmen kuukauden ajanjaksolla Nyhtökauraa oli haettu keskimäärin eniten sunnuntaina (indeksi noin 50) ja vähiten tiistaina (indeksi noin 20). Tuotteen hakuindeksien keskiarvoissa ei ollut kuitenkaan tilastollisesti merkitsevää eroa eri viikonpäivien välillä 5 %:n riskitasolla.

Nyhtökauran seitsemän päivän trendissä on eroja eri viikonpäivien ja vuorokaudenaikojen välillä. Nyhtökauraa haettiin seitsemän vuorokauden keskiarvojen perusteella melko tasaisesti pitkin päivää alkaen aamuyhdeksästä iltapäivän hakuhiippuun klo 17 (indeksi > 40) ja aina klo 18 asti. Nyhtökauraa ei haettu yöllä klo 3–4 ollenkaan. Nyhtökauran hakuindeksien keskiarvojen välillä oli tilastollisesti merkitseviä eroja eri kellonaikoina. Esimerkiksi klo 17 ja klo 19–9 ja 13 eroavat tilastollisesti merkitsevästi 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Nyhtökaura tuotiin markkinoille toukokuussa 2016. Vuoden 2015 kartalla näkyy, että nyhtökauraa oli haettu tuolloin jo Uudellamaalla. Nyhtökaura herätti eniten kiinnostusta suhteessa muihin vuosiin uutuuttaan vuonna 2016 Keski-Pohjanmaalla, Etelä-Savossa, Etelä-Karjalassa, Etelä-Pohjanmaalla, Satakunnassa ja Lapissa.

Nyhtökauran hakeminen ikäluokittain

Nyhtökauraa hakivat eniten 35–44- (34 %) ja 25–34-vuotiaat (32 %) ja lähes yhtä paljon 18–24-vuotiaat (28 %). Hakuja eivät tehneet ollenkaan tai tekivät hyvin vähän alle 18- tai yli 55-vuotiaat.

Nyhtökauran hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 16.

6.2.8 Kasviproteiinit/soijarouhe

Trendi ja kausiluonteisuus

Soijarouheen hakeminen Google Trends -indekseillä viiden vuoden ajanjaksolla näyttää trendinä hieman nousevalta. Viiden erillisen vuoden hauista nähdään, että soijarouhetta haetaan melko tasaisesti ympäri vuoden, joten tuote ei ole kausiluonteinen.

Kuukausi- ja viikkokohtaiset erot hauissa

Soijarouhetta on haettu keskimäärin eniten tammikuussa, mutta melko tasaisesti ympäri vuoden. Tuotteen hakuindeksien keskiarvoissa oli merkitseviä eroja esimerkiksi tammi-kuun ja maalisi–heinäkuun sekä marras–joulukuun välillä 5 %:n riskillä. Eniten hakuja on tehty tammikuun viikoilla 1–4 (indeksi > 50). Hauissa näkyi nousua heinäkuun puolestavälistä (vk 29) elokuun kolmannelle viikolle (vk 34), jonka kanssa saman hakumäärän saavuttivat myös lokakuiset viikot 42–43 (indeksi noin 50). Soijarouheen hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitseviä eroja esimerkiksi viikkojen 3–4 ja 23, 27, 48 ja 50–51 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Kolmen kuukauden ajanjaksolla soijarouhetta haettiin eniten viikon alkupuolella tiistaina (indeksi < 40) ja lähes yhtä paljon myös keskiviikkona ja torstaina (indeksi lähes 40). Tuotetta haettiin vähiten perjantaina (indeksi < 20). Soijarouheen hakuindeksien keskiarvoissa ei ollut eri päivien välillä tilastollisesti merkitseviä eroja 5 %:n riskitasolla.

Soijarouheen seitsemän päivän trendissä näkyy kolme selkeää hakuhiippua sekä eroja viikonpäivien ja kellonaikojen välillä. Soijarouhetta haettiin seitsemän vuorokauden keskiarvojen perusteella eniten iltapäivällä klo 15 (indeksi > 40), ja kiinnostus säilyi tasaisen korkeana klo 16–18 (indeksi < 40). Kaikkiaan tuotetta haettiin aamuneljästä nousujohteisesti, mutta polveillen iltapäivään asti. Tuotetta ei haettu ollenkaan yöllä klo 1–2. Soijarouheen hakuindeksien keskiarvojen välillä oli tilastollisesti merkitseviä eroja eri kellonaikoina. Esimerkiksi klo 15 ja klo 3, 5, 8, 21, 22 tai 24 (0) eroavat tilastollisesti merkitsevästi 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Soijarouhetta haettiin ylipäätään vain seitsemässä maakunnassa, jotka pysyivät samoina eri vuosina. Hakuintensiteetti vaihteli eri maakuntien ja eri vuosien välillä, ollen suurinta vuonna 2019 Pohjois-Pohjanmaalla, Pirkanmaalla ja Lapissa.

Soijarouheen hakeminen ikäluokittain

Soijarouhetta hakivat eniten 18–24-vuotiaat (43 %), mutta vähemmän enää 25–34- (28 %) ja 35–44-(23 %) vuotiaiden ikäryhmät. Soijarouhetta eivät hakeneet ollenkaan tai hakivat hyvin vähän alle 18- ja yli 65-vuotiaat.

Soijarouheen hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 17.

6.2.9 Kasvisproteiinit/Härkis

Trendi ja kausiluonteisuus

Härkiksen hakeminen ajanjaksolla 20.8.2016–31.12.2019 näyttää Google Trends -indeksien trendinä hieman laskevalta. Viiden erillisen vuoden hauista nähdään, että Härkistä haetaan melko tasaisesti ympäri vuoden. Suuremmat piikit loppuvuonna 2016 ja alkuvuonna 2017 ajoittuvat lanseerauksen lähettyville, sillä Härkis tuotiin markkinoille syyskuussa 2016.

Kuukausi- ja viikkokohtaiset erot hauissa

Härkistä haettiin keskimäärin eniten ajanjaksolla 20.8.2016–31.12.2019 helmikuussa, mutta sitä haettiin pitkin vuotta painottuen kevääseen ja syksyyn. Tuotteen hakuindeksien keskiarvoissa oli merkitseviä eroja eri kuukausien välillä. Esimerkiksi tammikuu ja touko–elokuu tai marras–joulukuu erosivat merkitsevästi 5 %:n riskillä. Eniten Härkistä haettiin ajanjakson aikana keskimäärin syyskuun viikolla 39 (indeksi noin 50) ja alkuvuodesta viikkoina 3–9 (indeksi suurempi tai yhtä suuri kuin 40). Tuotteen hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitsevä ero viikon 2 ja 18 tai 21 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Härkistä oli haettu kolmen kuukauden ajanjaksolla keskimäärin eniten sunnuntaina (indeksi < 50) ja vähiten perjantaina (indeksi < 20). Tuotteen hakuindeksien keskiarvoissa ei ollut tilastollisesti merkitsevää eroa eri viikonpäivien välillä 5 %:n riskitasolla.

Härkiksen seitsemän päivän trendissä on yhden päivän kohdalle osuva huippu, ja siinä näkyy eroja viikonpäivien ja kellonaikojen välillä. Härkistä haettiin seitsemän vuorokauden keskiarvojen perusteella nousujohteisesti ja polveillen aamuseitsemästä iltapäivän hakuhiippuun klo 16–17 asti (indeksi noin 35). Tuotteen hakuindeksien keskiarvojen välillä oli

tilastollisesti merkitseviä eroja eri kellonaikoina. Esimerkiksi klo 16–17 ja klo 1–6, 8, 18–19 tai 21–24 (0) eroavat tilastollisesti merkitsevästi 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Härkis tuotiin markkinoille vuonna 2016, jolloin sitä haettiin kaikissa maakunnissa. Eniten sitä haettiin kuitenkin lähes kaikissa maakunnissa vasta seuraavana vuonna. Vuonna 2018 Härkistä haettiin enää hyvin vähän, jos ollenkaan, Kainuussa. Härkistä haettiin eniten Keski-Pohjanmaalla vuonna 2019.

Härkiksen hakeminen ikäluokittain

Härkistä hakivat eniten 35–44-vuotiaat (46 %). 18–24- ja 25–34-vuotiaiden osuudet hauista olivat 23 ja 28 %. Alle 18- ja yli 55-vuotiaat eivät hakeneet ollenkaan tai hakivat vain vähän Härkistä.

Härkiksen hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 18.

6.2.10 Sesonkituotteet/porkkanalaatikko

Trendi ja kausiluonteisuus

Porkkanalaatikon trendi on lievästi laskeva Google Trendsin viiden vuoden hakuindeksien perusteella. Viiden eri vuoden (2015–2019) indekseistä näkyy selkeästi tuotteen kausiluonteisuus, sillä hakupiikki osuu kaikkina vuosina samaan ajankohtaan joulun seutuville eikä hakuja ole juurikaan muina ajankohtina.

Kuukausi- ja viikkokohtaiset erot hauissa

Porkkanalaatikkaa haettiin keskimääräisesti viiden vuoden aikana (2015–2019) ylivoimaisesti eniten joulukuussa. Ruoan hakuindeksien keskiarvoissa oli merkitsevät erot joului- ja marraskuun sekä kaikkien muiden kuukausien välillä sekä myös marras- ja joulukuun välillä 5 %:n riskillä. Porkkanalaatikon hakujen määrä lähti loivaan yhtenäiseen nousuun viikosta 36 alkaen ja nousi jyrkästi hakuhiippuun viikolla 50 (indeksi 80). Viikolla 51 ruokaa haettiin toiseksi eniten (indeksi > 40), mutta viikolla 52 haut romahtavat lähes nolnaan. Porkkanalaatikon hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitsevä ero viikon 50 ja kaikkien muiden paitsi viikon 51 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Porkkanalaatikkaa oli haettu keskimäärin kolmen kuukauden ajanjaksolla eniten sunnuntaina ja maanantaina (indeksit > 10), mutta hakemisen intensiteetti oli ylipäättään alhainen. Porkkanalaatikon hakuindeksien keskiarvoissa ei ollut tilastollisesti merkitsevää eroa eri viikonpäivien välillä 5 %:n riskitasolla.

Porkkanalaatikon seitsemän päivän hakutrendissä näkyy vuorokaudenaikoihin liittyvää säännöllisyyttä ja eroja eri päivien välillä. Porkkanalaatikon hakuindeksien keskiarvoja tarkasteltiin joulun aikana päivittäin ja kellonajoittain takautuvasti 25.12.2019 alkaen seitsemän vuorokauden ajan, koska tuote on selkeästi sesonkiluoteinen. 19.12. haut lähtivät nousemaan, jolloin hakuhuipun indeksit olivat lähes 30. Seuraavana päivänä suurimmat hakuhuiput jäivät vielä alle 40:n. 21.12. hakuhuiput nousivat noin indeksiin 60–75 klo 8–14. Eniten hakuja oli klo 11. Kaksi päivää ennen jouluaattoa 22.12. indeksit saavuttivat viikon jakson korkeimman huipun (indeksi 100) klo 10–11. Indeksit olivat erittäin korkeita > 80 myös klo 12–14 ja 16, kuten myös aatonaattona klo 10–18. Jouluaattona hakuintensiteetti hieman laski ollen suurimmillaan klo 8–13 (> 60). Joulupäivänä haut tippuivat reilusti, jolloin indeksit olivat enää < 20. Porkkanalaatikon hakuindeksien keskiarvot eroavat tilastollisesti merkitsevästi klo 10–11 ja klo 23–4 välillä 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Porkkanalaatikkoo haetaan melko tasaisesti ympäri maan. Eri vuosien välillä on maakuntakohtaista vaihtelua.

Porkkanalaatikon hakeminen ikäryhmittäin

Porkkanalaatikkoo hakevat eniten 25–34- (39 %) ja 35–44-vuotiaat (40 %). Alle 18- ja yli 65-vuotiaat eivät hae ollenkaan porkkanalaatikkoo tai hakujen määrä on hyvin pieni.

Porkkanalaatikon hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 19.

6.2.11 Sesonkituotteet/lanttulaatikko

Trendi ja kausiluonteisuus

Lanttulaatikon trendi on lievästi laskeva Google Trendsin viiden vuoden hakuindeksien perusteella. Viiden eri vuoden (2015–2019) indekseistä näkyy selkeästi tuotteen kausiluonteisuus, sillä hakupiikki osuu kaikkina vuosina samaan ajankohtaan joulun seutuville eikä hakuja ole juurikaan muina ajankohtina.

Kuukausi- ja viikkokohtaiset erot hauissa

Lanttulaatikkoo haettiin keskimääräisesti viiden vuoden aikana (2015–2019) ylivoimaisesti eniten joulukuussa. Ruoan hakuindeksien keskiarvoissa oli merkitsevät erot joului- ja marraskuun sekä kaikkien muiden kuukausien välillä sekä myös marras- ja joulukuun välillä 5 %:n riskillä. Lanttulaatikon haut alkoivat viikolla 36 ja lähtivät lievään nousuun viikosta 43 alkaen nousten hakuhuippuun viikolla 50 (indeksi < 80). Viikolla 51 ruokaa haettiin toiseksi

eniten (indeksi < 50), mutta viikolla 52 haut romahtavat lähes nollaan. Lanttulaatikon hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitsevä ero viikon 50 ja kaikkien muiden paitsi viikon 51 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Lanttulaatikkoa haettiin keskimäärin kolmen kuukauden ajanjaksolla eniten maanantaina ja sunnuntaina (indeksit > 10), mutta hakemisen intensiteetti oli ylipäättään alhainen. Porkkanalaatikon hakuindeksien keskiarvoissa ei ollut tilastollisesti merkitsevää eroa eri viikonpäivien välillä 5 %:n riskitasolla.

Lanttulaatikon seitsemän päivän hakutrendissä näkyy vuorokaudenaikoihin liittyvää säännöllisyyttä ja eroja eri päivien välillä. Lanttulaatikon hakuindeksien keskiarvoja tarkasteltiin joulun aikana päivittäin ja kellonajoittain takautuvasti 25.12.2019 alkaen seitsemän vuorokauden ajan, koska tuote on selkeästi sesonkiluoteinen. 19.12. haut lähtivät nousemaan, jolloin hakuhiipun indeksit olivat noin 30. Seuraavana päivänä suurimmat hakuhiiput ylsivät 40:n. 21.12. hakuhiiput nousivat noin indeksiin 70 klo 8–16. Kaksi päivää ennen jouluaattoon 22.12. indeksit saavuttivat viikon jakson korkeimman huipun (indeksi 100) klo 12. Indeksit olivat erittäin korkeita > 80 myös klo 9–11, 13–15 ja 17–18, kuten myös aattona klo 9–18. Jouluaattona hakuintensiteetti hieman laski ollen suurimmillaan klo 8–14 (indeksi 60–80). Hakupiikki oli klo 9. Joulupäivänä haut tippuivat reilusti, jolloin indeksit olivat enää < 20. Lanttulaatikon hakuindeksien keskiarvot eroavat tilastollisesti merkitsevästi klo 7–19 ja klo 24 (0)–4 välillä 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Lanttulaatikkoa haetaan melko tasaisesti ympäri maan. Eri vuosien välillä on maakunta-kohtaista vaihtelua.

Lanttulaatikon hakemisen ikäjakauma

Lanttulaatikkoa hakivat eniten 25–34-vuotiaat (43 %) ja lähes yhtä paljon ikäluokkaan 35–44 kuuluvat henkilöt (36 %). Alle 18- tai yli 65-vuotiaat eivät hakeneet lanttulaatikkoa tai hakujen määrä oli hyvin pieni.

Lanttulaatikon hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 20.

6.2.12 Sesonkituotteet/joulukinkku

Trendi ja kausiluonteisuus

Joulukinkun trendi on lievästi laskeva Google Trendsin viiden vuoden hakuindeksien perusteella. Viiden eri vuoden (2015–2019) indekseistä näkyy selkeästi tuotteen kausiluonteisuus, sillä hakupiikki osuu kaikkina vuosina samaan ajankohtaan joulun seutuville eikä hakuja ole juurikaan muina ajankohtina.

Kuukausikohtaiset erot hauissa

Joulukinkkua haettiin keskimääräisesti viiden vuoden aikana (2015–2019) lähes yksinomaan joulukuussa. Ruoan hakuindeksien keskiarvoissa oli merkitsevät erot joulu- ja marraskuun sekä kaikkien muiden kuukausien välillä sekä myös marras- ja joulukuun välillä 5 %:n riskillä. Joulukinkun haut lähtivät lievään nousuun viikosta 41 alkaen ja jyrkkenivät viikolla 50 nousten hakuhuippuun viikolla 51 (indeksi < 60), jonka jälkeen haut romahtavat hyvin pieniksi. Joulukinkun hakuindeksien keskiarvoissa eri viikkoina oli tilastollisesti merkitsevä ero viikon 51 ja kaikkien muiden paitsi viikkojen 49–50 välillä 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Joulukinkkua haettiin keskimäärin kolmen kuukauden ajanjaksolla eniten maanantaina ja sunnuntaina (indeksit < 10, jouluaatto oli tiistaina), mutta hakemisen intensiteetti oli ylipäättään alhainen. Ruoan hakuindeksien keskiarvoissa ei ollut tilastollisesti merkitsevää eroa eri viikonpäivien välillä 5 %:n riskitasolla.

Joulukinkun seitsemän päivän hakutrendissä on yksi selkeä huippu. Eri päivien ja kellonaikojen väliset erot näyttävät pieniltä. Joulukinkun hakuindeksien keskiarvoja tarkasteltiin joulun aikana päivittäin ja kellonajoittain takautuvasti 25.12.2019 alkaen seitsemän vuorokauden ajan, koska tuote on selkeästi sesonkiluoteinen. 19.12. haut lähtivät hitaasti nousemaan 22.12. asti, jolloin haut ylittivät indeksin 20. Joulun aattonaaton hakuindeksit ylittivät 40 rajapyykin klo 15, jonka jälkeen haut jatkoivat nousua välillä hieman polveilleen aina jouluaattoamuun klo 5 asti, jolloin saavutettiin hakuhuippu (indeksi 100). Hakeminen oli erittäin intensiivistä myös jouluaattona klo 3 ja 4 (indeksit noin 85–95). Klo 6 jälkeen haut alkoivat laskea melko jyrkästi ja painuivat lähelle nollaa klo 18 jälkeen. Joulukinkun hakuindeksien keskiarvot eivät eroa tilastollisesti merkitsevästi eri kellonaikojen välillä 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Joulukinkkua haetaan tasaisesti ympäri maan. Eri vuosien välillä on maakuntakohtaista vaihtelua. Hakuintensiteetti kasvoi vuonna 2019 koko maassa verrattuna aikaisempiin vuosiin.

Joulukinkun hakeminen ikäluokittain

Joulukinkkua hakivat eniten 25–34-vuotiaat (39 %) ja lähes yhtä paljon 18–24-vuotiaiden ryhmä (35 %). Alle 18- ja yli 65-vuotiaat eivät hakeneet ollenkaan joulukinkkua tai hakujen määrä oli hyvin pieni.

Joulukinkun hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 21.

6.2.13 Punaviini

Trendi ja kausiluonteisuus

Punaviinin hakutrendi oli Suomessa lievästi nouseva viiden vuoden tarkastelujaksona. Yksittäisiä vuosia 2015–2019 verrattaessa huomataan, että punaviinin suurimmat hakupiikit osuivat kaikkina vuosina joulun seudulle. Myös pääsiäinen näkyy piikkinä hakuindekseissä, mutta joulua pienempänä. Punaviiniä haetaan kuitenkin ympäri vuoden.

Kuukausi- ja viikkokohtaiset erot hauissa

Punaviiniä haettiin keskimäärin eniten vuosina 2015–2019 joulukuussa, mutta sitä haetaan muina kuukausina melko tasaisesti ympäri vuoden. Juoman hakuindeksien keskiarvoissa oli merkitsevät erot joulukuun ja kaikkien muiden kuukausien välillä sekä esimerkiksi huhtikuun ja loka- ja marraskuun välillä 5 %:n riskillä. Punaviinin suurimmat hakupiikit osuvat joulukuun viikkoihin 50 ja 51 (indeksit > 70) ennen joulua. Juoman hakuindeksit ovat noin 20 vuoden alusta syyskuun loppuun. Juhannusta edeltävä viikko 24 näkyy pienenä nousuna hauissa (indeksi noin 30). Lokakuun alkupuolella viikolla 41 alkaa punaviinin hakuintensiteetti hiljalleen nousta viikkoon 49 asti, josta se ponkaisee joulua edeltäviin huippulukemiin. Punaviinin hakuindeksien keskiarvoissa oli tilastollisesti merkitseviä eroja eri viikkojen välillä. Esimerkiksi viikot 50 ja 51 eroavat merkitsevästi muista viikoista lukuun ottamatta viikkoja 48 ja 49 tai viikko 24 eroaa viikoista 22, 23 ja 25 tilastollisesti merkitsevästi 5 % riskillä.

Erot hakemisessa eri viikonpäivinä ja kellonaikoina

Punaviiniä haettiin keskimäärin kolmen kuukauden ajanjaksolla eniten perjantaina ja lauantaina (indeksit > 20). Juoman hakuindeksien keskiarvoissa oli tilastollisesti merkitsevä ero keskiviikon ja lauantain välillä 5 %:n riskitasolla.

Punaviinin hakutrendissä seitsemän päivän aikana joulun seudulla, oli eroja viikonpäivien ja kellonaikojen välillä. Punaviinin hakuindeksien keskiarvoja tarkasteltiin joulun aikana päivittäin ja kellonajoittain takautuvasti 25.12.2019 alkaen seitsemän vuorokauden ajan, koska tuotetta haetaan selkeästi eniten ennen joulua. 19.12. päivittäisten hakuhuippujen trendi oli nouseva aina jouluaattoon klo 15 asti, joka oli suurin hakuhuippu (indeksi 100). Poikkeuksen tekee 22.12., joka on viikonpäivältään sunnuntai. Tyypillisesti punaviiniä haettiin iltapäivällä ja illan suussa ja hakuja oli vähemmän yöllä ja aamulla. 20.12. punaviiniä haettiin eniten klo 17–21 ja 21.12. klo 14 ja 16–18 (indeksit > 60). Aatonaattona punaviiniä haettiin eniten klo 12–18 ja jouluaattona 14–17 (indeksit > 80), jonka jälkeen haut hiljalleen putosivat lähelle nollaa aamuyön tunteina. Punaviinin hakeminen kiinnosti myös joulupäivänä, sillä klo 14 hakuindeksi ylitti 40 rajapyykin. Punaviinin hakuindeksien keskiarvot eroavat tilastollisesti merkitsevästi esimerkiksi iltapäivän–illansuun klo 16–18 ja yön–aamun klo 23–8 välillä 5 %:n riskillä.

Aluekohtaiset erot eri vuosina

Punaviiniä haettiin melko tasaisesti ympäri maan eri vuosina. Eri vuosien välillä on maakuntakohtaista vaihtelua. Esimerkiksi Keski-Pohjanmaan ja Etelä-Karjalan hakuintensiteetti oli vuonna 2019 muita maakuntia suurempaa.

Punaviinin hakemisen ikäjakauma

Punaviiniä hakivat eniten 18–24-vuotiaat (59 %). Muut ikäluokat, kuten 25–34-vuotiaat (27 %), hakivat sitä huomattavasti vähemmän. Alle 18- ja yli 55-vuotiaat eivät hakeneet punaviiniä ollenkaan tai hakumäärä oli hyvin pieni.

Punaviinin hakuajankohtia kuvaavat graafit ja kartat ovat liitteessä 22.

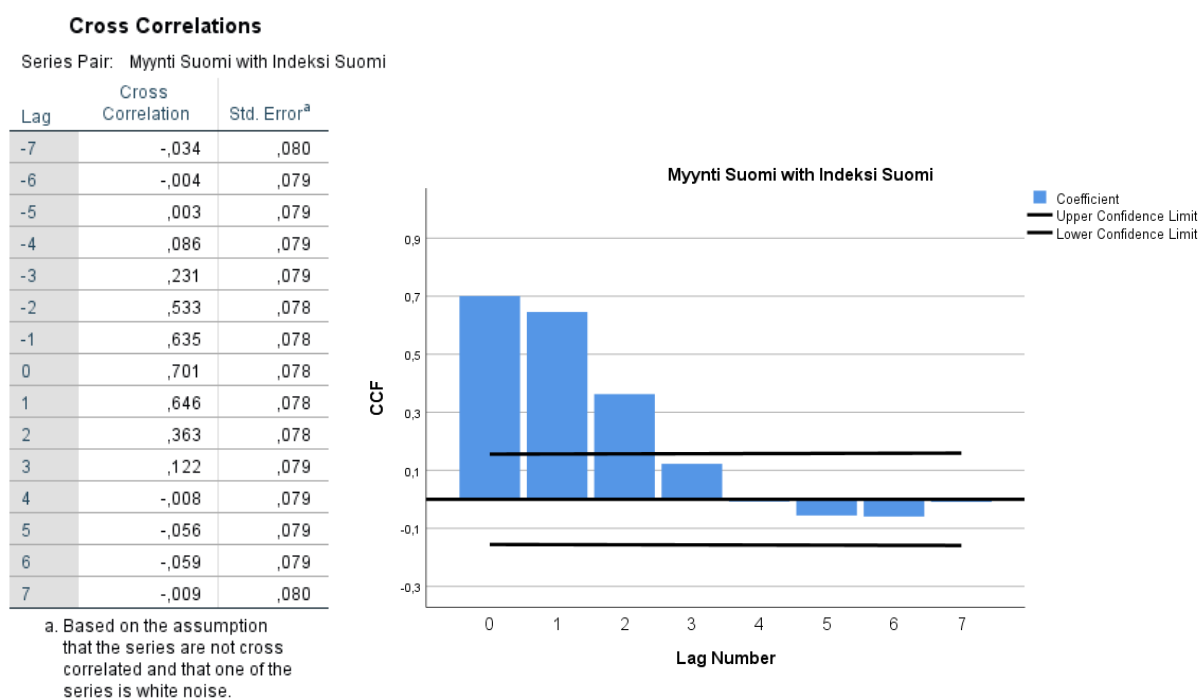
6.3 Punaviinin Google-hakujen ja ostamisen välinen viive

SPSS-ohjelman kahden aikasarjamuuttujan ristikorrelaatioanalyysillä saatiin selville punaviinin hakemisen ja myynnin viiveluvut sekä korrelaatiot Suomessa ja maakunnissa 95 %:n luottamusvälillä. Taulukosta 9 nähdään myyntien ja indeksien ristikorrelaatiot Suomessa eri viiveillä, ja kuviosta 9 viiveiden luottamusvälit. Maakuntien tulokset on koottu liitteeseen 8 ja yhteenveto Excel-tilaukiksi liitteeseen 9. Viive tarkoittaa aikaa Google-hausta myyntitapahtumaan. Myynnin ja indeksin aikajänne on kuukausi, koska tällä tarkkuudella olivat myös alkuperäiset datat.

Viivelukuja tulkitaan seuraavasti:

- 0 haun jälkeen myynti saman kuukauden aikana
- 1 haun jälkeen myynti seuraavan kuukauden aikana
- 2 haun jälkeen myynti sitä seuraavan kuukauden aikana.

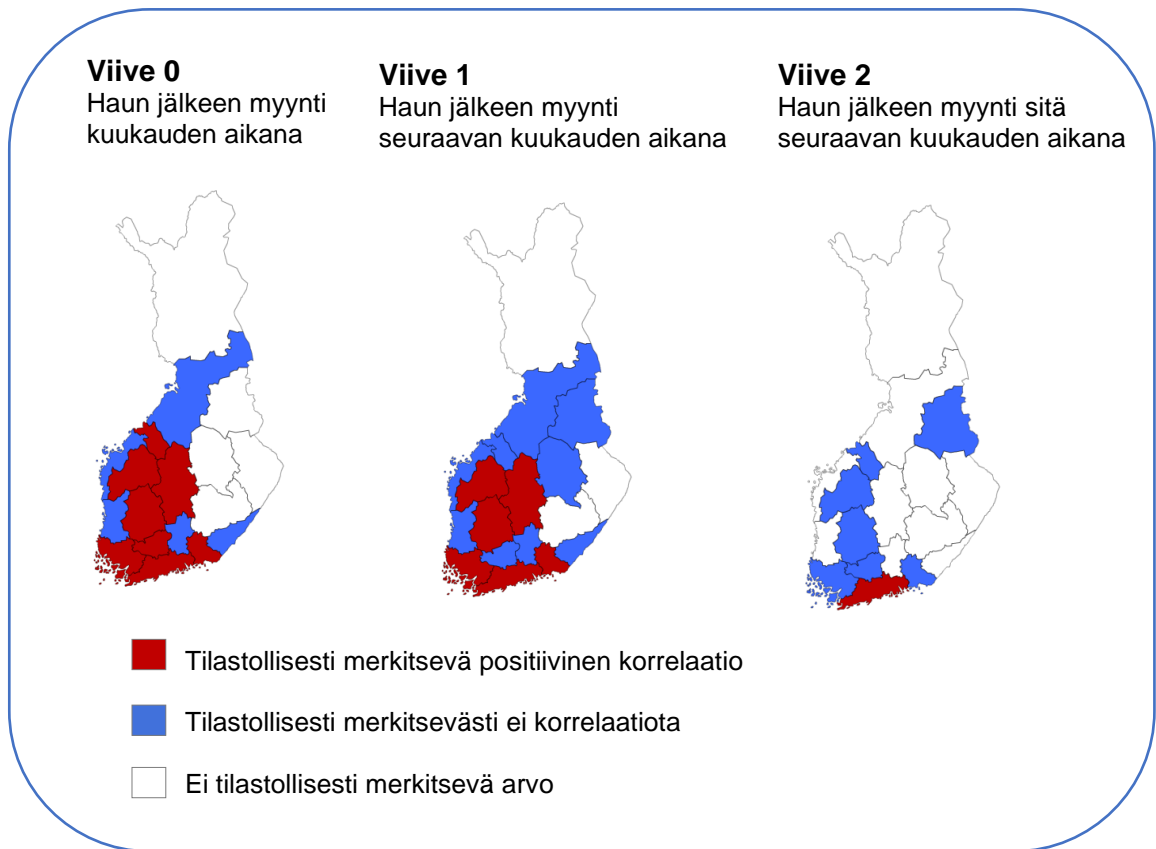
Taulukko 9 ja kuvio 9. SPSS:n ristikorrrelaatioanalyysin punaviinin hakemisen ja myynnin korrelaatiot ja viiveet keskivirheineen Suomessa (vasen) ja luottamusvälit (oikea)



Voimakas positiivinen korrelaatio ($r = 0,701$, luottamusväli 95 % = $[0,623]$, $[0,779]$) punaviinin hakemisen ja myynnin välillä Suomessa osuu viiveluvun arvolle 0. Hakemisen ja myynnin välillä on vielä kohtalainen positiivinen korrelaatio yhden ($r = 0,646$, luottamusväli 95 % = $[0,568]$, $[0,724]$) ja kahden ($r = 0,363$, luottamusväli 95 % = $[0,285]$, $[0,441]$) kuukauden viiveellä hakemisesta, jonka jälkeen ei ole enää havaittavissa riippuvuutta 95 % virhemarginaalilla. Kuvio 9 poistettiin negatiiviset arvot, koska kiinnostus kohdistuu Google-hakujen vaikutukselle myyntiin. Suurin korrelaatio osuu viiveluvun arvolle 0. Tämä tarkoittaa, että punaviiniä myydään eniten hakua seuraavan kuukauden aikana. Luottamusrajoista nähdään, että 5 % riskillä viiveluvut 0–2 ovat tilastollisesti merkitseviä. Punaviiniä myydään myös haun jälkeisenä kuukautena ja sitä seuraavana kuukautena, mutta vähemmän kuin saman kuukauden aikana.

Spatiotemporaalisella analyysillä kuvattiin maakuntien välisiä eroja punaviinin Google-hakujen ja myynnin välillä viivelukujen arvoja 0, 1 ja 2 vastaavilla kartoilla (kuvio 10). Viiveellä 0 haun ja myynnin välillä on kohtalainen (0,3–0,7) tilastollisesti merkitsevä positiivinen korrelaatio kahdeksassa maakunnassa. Sen sijaan viiveellä 1 korrelaatio on havaittavissa enää kuudessa maakunnassa ja viiveellä 2 ainoastaan Uudellamaalla. Punainen väri kuvaa tilastollisesti merkitsevää positiivista korrelaatiota Google Trends -hakuindek-

sien ja punaviinin myynnin välillä. Sininen väri tarkoittaa, että muuttujien välillä ei ole korrelaatiota, mutta tulos on tilastollisesti merkitsevää. Valkoinen väri kuvaa arvoja, jotka eivät ole tilastollisesti merkitseviä.



Kuvio 10. Punaviinin hakemisen ja myynnin väliset erot eri maakunnissa viiveluvuilla 0, 1 ja 2

6.4 Punaviinin myynnin ennustemalli

Kuvaileva tilastoanalyysi

Tehtiin kuvaileva tilastoanalyysi datoilte (taulukko 10). Sekä punaviinin myynnin että Google Trends -indeksien havaintojen lukumäärä (n) on 165.

Myyntin arvot vaihtelevat 1 381 000 (minimi)–3 438 000 (maksimi) välillä. Koska myynnin keskiarvo 1 897 224, on mediaania 1 769 000 suurempi, vaikuttaa jakauma jonkin verran vinolta oikealle. Keskihajonta 38 736 kertoo, kuinka paljon havainnot keskimäärin poikkeavat keskiarvosta. (Taanila 21.4.2019.) Prosenttipisteistä nähdään, että punaviinin myynnistä 25 % jää alle 1 614 000 litraa ja vain 10 % ylittää 2 355 000 litraa.

Indeksien arvot vaihtelevat 11 (minimi)–100 (maksimi) välillä. Koska indeksien keskiarvo 25,75, on mediaania 23,00 suurempi, vaikuttaa jakauma jonkin verran vinolta oikealle.

Keskihajonta 0,954 kertoo, kuinka paljon havainnot keskimäärin poikkeavat keskiarvosta. (Taanila 21.4.2019.) Indekseistä 90 % on alle 37, joten vain 10 % on tuota indeksilukua suurempia.

Taulukko 10. SPSS:n punaviinin hakemisen Google Trends -indeksien ja myynnin tunnuslukuyhteenveto Suomessa

Statistics			
		Myynti Suomi	Indeksi Suomi
N	Valid	165	165
	Missing	0	0
Mean		1897224,24	25,75
Std. Error of Mean		38735,545	,954
Median		1769000,00	23,00
Mode		1502000 ^a	18 ^a
Std. Deviation		497567,085	12,250
Variance		2,476E+11	150,066
Range		2057000	89
Minimum		1381000	11
Maximum		3438000	100
Percentiles	10	1502000,00	16,00
	25	1614000,00	19,00
	50	1769000,00	23,00
	75	1982000,00	28,50
	90	2355000,00	37,00

a. Multiple modes exist. The smallest value is shown

Pearsonin korrelaatio

Google Trends -indeksien ja punaviinin myynnin välillä on voimakas positiivinen riippuvuus Suomessa ($r = 0,701$, $n = 165$, $p\text{-arvo} < 0,001$).

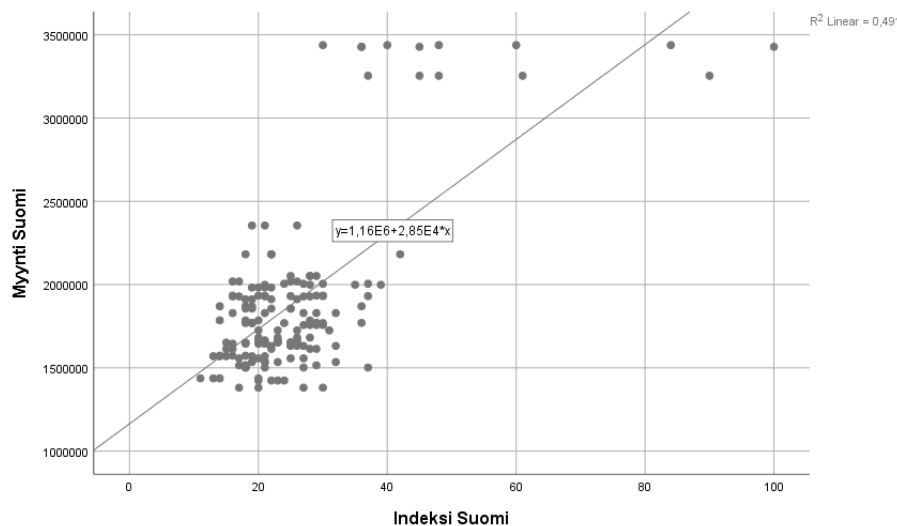
Lineaarinen regressioanalyysi

Lineaarisen regressioanalyysin punaviinin hakuindeksien ja myynnin riippuvuutta kuvaavasta hajontakuviosta (kuvio 11) nähdään, että datapisteiden hajonta on suurta, kun indeksi ylittää arvon 30. Regressiosuoran avulla arvioitiin Google Trends -hakuindeksien (selittävä muuttuja x = Indeksi Suomi) ja punaviinin myynnin (selitettävä muuttuja y = Myynti Suomi) riippuvuutta Suomessa. Suoran yhtälö on $y = a + bx$. Kertoimet-taulukosta 11 saatiin regressiosuoraan tarvittavat kertoimet sarakkeesta B: vakiotekijä $a = 1164384,796$ ja kulmakerroin (Indeksi Suomi) $b = 28458,110$. (Saaranen 2018, 69–71.) Vakiotekijä kertoo myynnin arvon, kun indeksi on arvoltaan 0. Toisin sanoen vakiotekijästä näkee, missä kohtaa kuviota regressiosuora leikkaa y-akselin. Vakiotekijän arvolla 1164384,796 myynti olisi 1164384,796 litraa, vaikka hakuja ei olisi ollenkaan. Yhtälön kul-

makertoimesta puolestaan nähdään, kuinka paljon punaviinin myynti muuttuu, kun hakuindeksi kasvaa yhdellä yksiköllä: indeksillä 1 punaviinin myynti on 1 192 843 litraa. Koska regressiosuoran kulmakerroin on positiivinen, suurentuva Google Trends -indeksi kasvattaa myös punaviinin myyntiä. Itseisarvoltaan huomattavasti nolasta poikkeavasta kulmakertoimesta näkee, että myynnin vaihtelua voidaan ennustaa ja muuttujalla x (indeksi) on merkitystä mallille. Mallin käyttöalue on selittävän muuttujan eli hakuindeksin arvo 0–100. (Taanila 2015, 11–12.), mutta ennustemäärät punaviinin myynnille laskettiin Google Trends -indeksin arvoille 10–100, sillä tältä alueelta oli käytettävissä havaintoja (taulukko 12):

y (= Myynti Suomi) = $1164384,796 + 28458,110x$ (x = Indeksi Suomi). Esimerkiksi hakuindeksillä 30 ennustettu myyntimäärä on 2 018 128 litraa.

Normitetun Google Trends -indeksin beta-kerroin 0,701 (taulukko 11) on positiivinen (beta 0,701, p-arvo < 0,01), jolloin indeksin kasvaessa yhdellä keskihajonnalla punaviinin myyntimäärä kasvaa 0,701 litralla (Taanila 2010, 10).



Kuvio 11. SPSS:n lineaarisen regressioanalyysin punaviinin hakuindeksien ja myynnin riippuvuutta kuvaava hajontakuvi

Taulukko 11. SPSS:n kertoimet-taulukko

Coefficients ^a						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	1164384,796	64695,903		17,998	,000
	Indeksi Suomi	28458,110	2269,966	,701	12,537	,000

a. Dependent Variable: Myynti Suomi

Taulukko 12. Lineaarisen regressiomallin ennustemäärät punaviinin myynnille litroina eri hakuindekseillä

Indeksi	10	20	30	40	50	60	70	80	90	100
Myynti litraa	1448966	1733547	2018128	2302709	2587290	2871871	3156452	3441034	3725615	4010196

Mallissa Indeksi Suomi -muuttujan kertoimen arvo on 28458,110 (taulukko 11) 95 %:n luottamusvälillä 23975,779–32940,441. Mallin selitysasteen ($R^2 = 0,491$) mukaan (taulukko 13) myynnin vaihtelusta voidaan selittää 49,1 % hakuintensiteetin vaihtelulla (Taanila 2015, 13).

Taulukko 13. SPSS:n mallin yhteenveto

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,701 ^a	,491	,488	356108,497

a. Predictors: (Constant), Indeksi Suomi

Punaviinin myynti ($n = 165$) oli keskimäärin 1 897 224 litraa (taulukko 14). Keskiarvon 95 % luottamusväli on 1 477 424–4 010 196 litraa eli 95 % varmuudella perusjoukon keskiarvo osuu edellä mainittuun haarukkaan (Taanila 19.4.2019b).

Taulukko 14. SPSS:n ennustettujen arvojen ja jäännösten perustunnusluvut

Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1477424,00	4010195,75	1897224,24	348615,827	165
Residual	-715334,875	1419871,875	,000	355021,140	165
Std. Predicted Value	-1,204	6,061	,000	1,000	165
Std. Residual	-2,009	3,987	,000	,997	165

a. Dependent Variable: Myynti Suomi

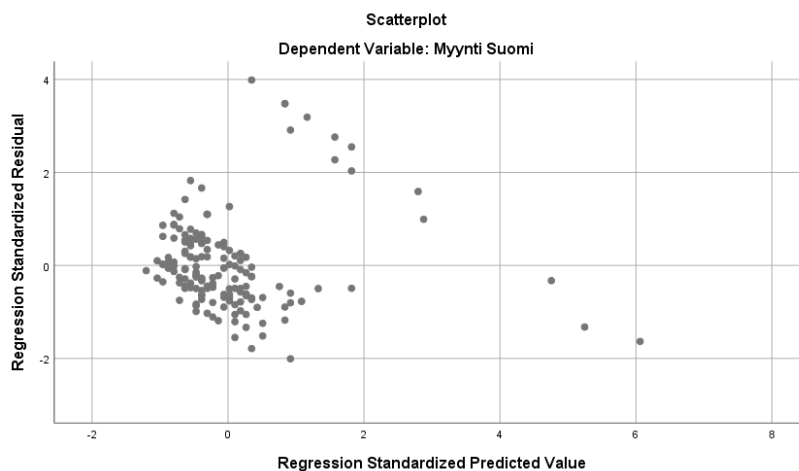
Mallin tilastollinen merkitsevyys

Edeltävyyssehtojen testaus

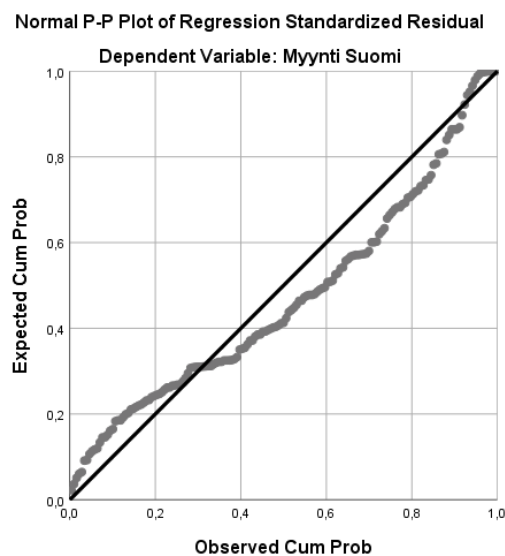
Lineaarinen riippuvuus ja jäännösten varianssien yhtäsuuruus

Kuviossa 12 pisteet jakautuvat x-akselin molemmin puolin melko tasaisesti, joten jäännöstermit viittaavat mallin toimivuuteen ja normaaliin jakaumaan. Lineaarisuuteen viittaa, että pisteet eivät jakaudu säännönmukaisen tasaisesti jäännöskuviossa. Kuvion 13 avulla verrattiin jäännösten normaalijakautuneisuutta vertaamalla havaintoja suoraan. Kuvion havainnot poikkeavat suorasta jonkin verran, mutta vakavia eroja ei ole, joten jakauman voidaan ajatella olevan normaalijakautunut. Tarkistettiin tulos vielä histogrammista (kuvio 14). Kuviosta nähdään, että virhetermit ovat normaalijakaumaa huipukkaampia. Lisäksi

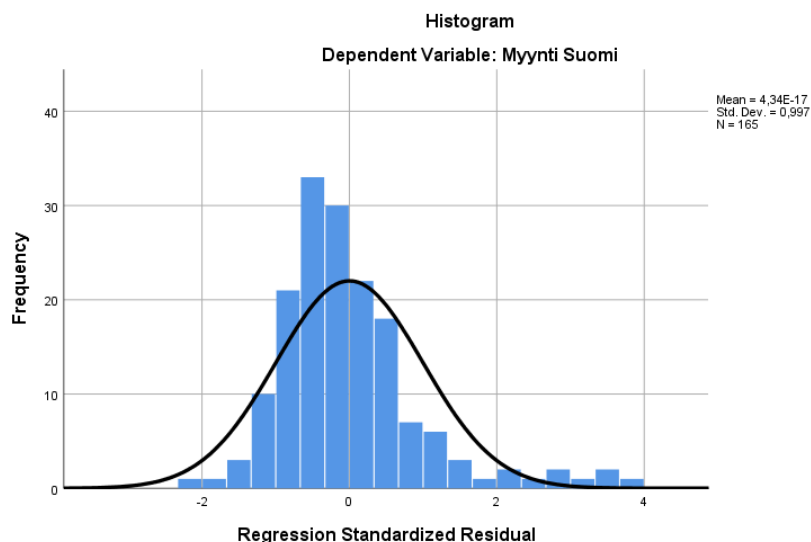
malli on hieman oikealle vino (Saaranen 2017, 60). Myynnin vinousluku (taulukko 15) on 2,130 (keskihajonta 0,189), joka viittaa vinouteen oikealle ja huipukkuusluku 4,018 (keskihajonta 0,376), joka kertoo normaalijakaumaa huipukkaammasta jakaumasta. Myyntijakauman vinous oikealle ja huipukkuus ovat tilastollisesti merkitseviä, koska $2,130/0,189 = 11,27 > 2$ ja $4,018/0,376 = 10,69 > 2$. Indeksien vinousluku on 3,296 (keskihajonta 0,189), joka viittaa vinouteen oikealle ja huipukkuusluku 14,798 (keskihajonta 0,376), joka kertoo normaalijakaumaa huipukkaammasta jakaumasta. Myyntijakauman vinous oikealle ja huipukkuus ovat tilastollisesti merkitseviä, koska $3,296/0,189 = 17,44 > 2$ ja $14,798/0,376 = 39,36 > 2$. (Saaranen 2018, 18–19.) Tämän mukaan jäännökset eivät ole normaalijakautuneita. Otokoko on kuitenkin riittävän suuri ($n > 30$) normaalijakaumaoletuksen kannalta.



Kuvio 12. SPSS:n mallin jäännöskuvio, jossa x-akselilla ovat ennustetut arvot ja y-akselilla jäännökset



Kuvio 13. SPSS:n jäännösten kumulatiivisen prosenttijakauman vertaaminen normeeratun normaalijakauman kertymäfunktioon (Saaranen 2017, 60)



Kuvio 14. SPSS:n histogrammi normeerattujen jäännöstermien jakauma verrattuna normaalijakaumaan

Taulukko 15. SPSS:n muuttujien vinous ja huipukkuus

Descriptive Statistics										
	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic Std. Error		Std. Deviation Statistic	Skewness Statistic Std. Error		Kurtosis Statistic Std. Error	
Myynti Suomi	165	1381000	3438000	1897224,24	38735,545	497567,085	2,130	,189	4,018	,376
Indeksi Suomi	165	11	100	25,75	,954	12,250	3,296	,189	14,798	,376
Valid N (listwise)	165									

Koska Durbin-Watsonin testin arvo 1,517 sijoittuu välille 1,5–2,5 ei jäännösten välillä ole autokorrelaatiota (taulukko 16) (Karadimitriou & Marshall 2020). Edellä tehtyjen edeltävyysehtojen testauksen perusteella voidaan sanoa, että merkitsevyystestaus on pätevä, koska edeltävyysehdot täyttyivät riittävästi. Merkitsevyystesteillä voidaan tarkastella mallin tilastollista merkitsevyyttä. (Taanila 2010, 11–15.)

Taulukko 16. SPSS:n jäännösten riippumattomuus Durbin-Watsonin testillä

Model Summary ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,701 ^a	,491	,488	356108,497	1,517

a. Predictors: (Constant), Indeksi Suomi

b. Dependent Variable: Myynti Suomi

Anova-taulukosta (taulukko 17) nähdään ($F = 157,171$, $p\text{-arvo} < 0,01$), että malli sopii aineistoon tilastollisesti merkitsevästi (Taanila 2010, 16; Nummenmaa 2011, 328).

Taulukko 17. SPSS:n ANOVA-tilaus

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,993E+13	1	1,993E+13	157,171	,000 ^b
	Residual	2,067E+13	163	1,268E+11		
	Total	4,060E+13	164			

a. Dependent Variable: Myynti Suomi

b. Predictors: (Constant), Indeksii Suomi

Selittävä muuttuja, Indeksii Suomi, ($t = 12,537$, $p\text{-arvo} < 0,01$) (tilaus 11) sopii malliin ja sitä voidaan pitää merkitsevästä selittäjänä (Taanila 2010, 16, 19).

Mallissa (tilaus 17) keskiarvomyyntimäärän 95 %:n luottamusväli on 23975,779–32 940,441 litraa (Taanila 2010, 18).

Taulukko 17. SPSS:n kertoimet-tilauksen luottamusvälit ennusteen keskiarvolle

Coefficients ^a			
		95,0% Confidence Interval for B	
Model		Lower Bound	Upper Bound
1	(Constant)	1036634,673	1292134,920
	Indeksi Suomi	23975,779	32940,441

a. Dependent Variable: Myynti Suomi

Jäännöstermien havaintojen lukumäärä $n = 165$. Virhearvio punaviinin myynnissä vaihtelee -715 335 (minimi)–1 419 872 (maksimi) litran välillä verrattuna keskiarvoon (tilaus 18). Keskihajonta 27 638 kertoo, kuinka paljon virhearvo keskimäärin poikkeaa keskiarvosta (Taanila 19.4.2019a). Jäännösvirheestä 25 % jää alle -220 754 litraa ja 25 % ylittää 154 266 litraa.

Tilauksesta 19 nähdään yksittäisten havaintojen toteutuneet ja ennustetut myynnit, jäännökset ja ennustetun keskiarvon luottamusvälin ala- ja ylärajat tietyllä päivämäärällä ja indeksillä. Jos esimerkiksi hakuindeksi on 30 13.10.2019, malli ennustaa, että punaviinin myynti on 2 018 128 litraa, jolloin virhearvio on 261 128 litraa. Virhearvio punaviinin myynnissä vaihtelee 1 960 168 (minimi)–2 076 088 (maksimi) litran välillä verrattuna keskiarvoon.

Taulukko 18. SPSS:n punaviinin myynnin virhetermien tunnuslukuyhteenveto

Statistics		
RES_1		
N	Valid	165
	Missing	0
Std. Error of Mean		27638,35826
Median		-78379,4319
Mode		391536,7877
Std. Deviation		355021,1399
Minimum		-715334,859
Maximum		1419871,909
Percentiles	25	-220753,761
	50	-78379,4319
	75	154265,8426

Taulukko 19. SPSS:n data-aineistoon lasketut ennusteiden luottamusvälit

Aika	MyyntiSuomi	IndeksiSuomi	PRE_1	RES_1	LMCI_1	UMCI_1
2019-10-27	1757000	27	1932753,76134	-175753,76134	1877725,93368	1987781,58901
2019-10-20	1757000	29	1989669,98097	-232669,98097	1933024,04853	2046315,91341
2019-10-13	1757000	30	2018128,09078	-261128,09078	1960167,88500	2076088,29656
2019-10-06	1757000	28	1961211,87116	-204211,87116	1905549,31238	2016874,42993
2019-09-29	1535000	23	1818921,32209	-283921,32209	1762806,67972	1875035,96446
2019-09-22	1535000	21	1762005,10246	-227005,10246	1703265,47186	1820744,73307

Taulukon lyhenteiden selitykset (Taanila 2010, 17–18):

PRE_1 = Mallin antamat havaintoja vastaavat ennusteet

RES_1 = Mallin ennustaman ja toteutuneen havaintoarvon erotukset

LMCI = Ennustetun keskiarvon luottamusvälin alaraja

UMCI = Ennustetun keskiarvon luottamusvälin yläraja.

Epälineaarinen käyräsovitus

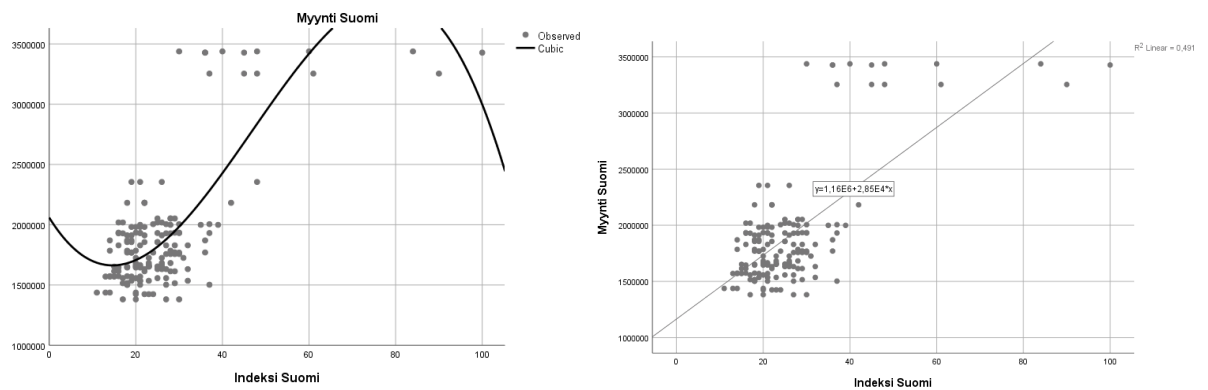
Kolmannen asteen käyrä, kuutiomalli, selittää parhaiten punaviinin myyntiä selitysasteella ($r^2 = 0,554$, p-arvo < 0,001), sillä muutokset Google Trends -indeksissä selittävät 55,4 % vaihtelusta punaviinin myynnissä (taulukko 20).

Verrattaessa kuutiomallin käyrää lineaariseen malliin nähdään, että kuutiomalli huomioi paremmin suurimpien indeksien vaikutuksen punaviinin myyntiin, koska pisteet osuvat tällöin lähemmäksi käyrää (kuvio 15). Verrattaessa koko mallin tilastollista merkitsevyyttä on lineaarinen malli ($F = 151,171$, p-arvo < 0,001) kuutiomallia ($F = 66,683$, p-arvo < 0,01) parempi (taulukko 20).

Taulukko 20. SPSS:n regressiomallin eri yhtälövaihtoehdot ja selitysasteet

Model Summary and Parameter Estimates									
Dependent Variable: Myynti Suomi									
Model Summary						Parameter Estimates			
Equation	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	,491	157,171	1	163	,000	1164384,796	28458,110		
Logarithmic	,442	129,328	1	163	,000	-1057529,582	930545,917		
Inverse	,317	75,800	1	163	,000	2796085,570	-20345008,1		
Quadratic	,499	80,711	2	162	,000	968984,060	40045,162	-126,780	
Cubic	,554	66,683	3	161	,000	2059298,983	-57831,361	2345,620	-16,734
Compound	,451	133,842	1	163	,000	1356962,502	1,012		
Power	,420	117,962	1	163	,000	521758,682	,398		
S	,313	74,303	1	163	,000	14,821	-8,875		
Growth	,451	133,842	1	163	,000	14,121	,012		
Exponential	,451	133,842	1	163	,000	1356962,502	,012		
Logistic	,451	133,842	1	163	,000	7,369E-7	,988		

The independent variable is Indeksi Suomi.



Kuvio 15. SPSS:n hajontakuviovertailu kuutiomallin (vasen) ja lineaarisen mallin (oikea) punaviinin myynnin ja Google Trends -indeksien välisestä riippuvuudesta Suomessa

7 Pohdinta

Pohdintaosuus koostuu tulosten tarkastelusta kohta 7.1., tutkimuksen luotettavuuden arvioinnista kohta 7.2. ja eettisten näkökohtien huomioimisesta kohta 7.3. Kohdassa 7.4. käydään läpi tutkimuksen johtopäätökset ja vastataan tutkimuskysymyksiin sekä tehdään kehittämis- ja jatkotutkimusehdotukset. Viimeisessä kohdassa 7.5. arvioidaan oman osaamisen kehittymistä.

7.1 Tulosten tarkastelu

Tulosten ja kehitysprosessin arviointi

Tutkimuksen tuloksia ja kehitysprosessia arvioidaan peilaten valittuihin työkaluihin, toteutukseen, mallien toimivuuteen ja käyttökelpoisuuteen kohdeyrityksen liiketoiminnan kannalta. Tutkimuksen onnistumista ja tulosten laatua mitataan toimeksiantajalta ja kohdeyritykseltä saadulla palautteella.

Tutkimuksen tulokset perustuvat teoriaan ja empiirisen tutkimuksen tuloksiin. Tutkimus täyttää kokonaisuutena asetetut tavoitteet. Dokumenttianalyysillä selvitettiin, millaista tietoa kuluttajista voi saada ja mitä ei voi saada yrityksen ulkoisista datalähteistä GDPR huomioiden. Empiirisessä osuudessa luotiin uudenlainen monialainen malli kuluttajan ostopäätösprosessin tiedonhaun ja ostopäätöksen ajankohdan selvittämiseksi. Mallissa hyödynnettiin hakukoneanalytiikkaa hakutrendien analysoinnissa ja Google Trends -massadataa. Lisäksi tehtiin myyntimäärille ennustemalli, joka hyödyntää Google Trends -hakudataa ja toteutuneita myyntejä. Kehitysprosessi onnistui kokonaisuudessaan, sillä kohdeyritys sai käyttöönsä uusia kuluttajaymmärrystä lisääviä malleja hyödyntää dataa erityisesti Google Trends -massadataa. Oma roolini kehitystyössä onnistui, koska pystyin liiketoiminnan taustatietojen perusteella suunnittelemaan ja toteuttamaan tutkimuksen, joka täytti asetetut tavoitteet.

Tulosten onnistuminen

Tutkimuksen tuloksia arvioidaan datapohjaista kuluttajaymmärrystä lisäävän ja kuluttajan ostopäätösprosessia selvittävän teoriaosuuden, tutkimusmenetelmien, datan ja työkalujen valinnan sekä lopputulosten perusteella.

Teoriaosuuden tavoitteena on jäsentää alaa ja sen kehitystä, selittää ilmiötä, hyödyntää aikaisemmin tehtyjen tutkimusten käsitteitä ja mittareita sekä peilata omaa työtä aikaisempiin tutkimuksiin (Kananen & Makkonen (toim.) 2015, 112). Tämän perusteella tutkimustulokset onnistuivat kokoamassani teoreettisessa viitekehyksessä kuluttajan ostopäätöspro-

sessista ja kuluttajaymmärryksen lisäämisestä datalla. Hyödynsin teoriaa aluksi lisätäk-
seni ymmärrystäni kohdeyrityksen liiketoiminnasta tutkimuskysymysten ja -tavoitteiden
asettamiseksi. Kanasen & Makkosen (toim.) (2015, 197) mukaan määrällisessä tutkimuk-
sessa on tärkeää perehtyä ilmiöitä selittäviin teorioihin ja niiden hyödyntämiseen. Näin ol-
len hyödynsin teoriaa edelleen tutkimusmenetelmien ja työkalujen valinnassa, datojen ke-
räämisessä, analysoinnissa ja mallintamisessa. Vertasin myös omia tuloksiani muihin tut-
kimuksiin.

Kaikkiaan empiirisen tutkimuksen tuloksissa näkyy uudenlainen, monialainen lähestymis-
tapa. Kaikissa tutkimusosuuksissa lähdin liikkeelle liiketoiminnan ymmärtämisestä. Pereh-
tyminen kuluttajan ostopäätösprosessiin ja ulkomainontaan liiketoimintana dokumentti-
analyysillä auttoi ymmärtämään, miten ulkomainonnan avulla voi vaikuttaa prosessin eri
vaiheissa. Tutkimuksessa tutkittiin Google Trends -hakuindeksien avulla vaihetta, jossa
kuluttaja etsii tietoa Googlestä tutkimuskohteeksi valituista päivittäiselintarvikkeista. Aluksi
varmistin Ubersuggest-työkalun hakutrendianalyysillä, että käytettävät hakutermit ovat sel-
laisia, joita kuluttajat oikeasti käyttävät. Samalla selvitin eri tuotenimikkeiden hakuvolyymit.
Tulokseksi saatiin monipuolinen kattaus tietoa, millä tarkkuudella dataa on saatavana
Google Trendsistä, milloin kuluttajat etsivät kyseisiä tuotteita lyhyellä ja pitkällä aikajän-
teellä ja miten haut eroavat eri maakunnissa. Tutkimuksessa havaittiin, että Google
Trends on käyttökelpoinen työkalu trendien ja kausiluonteisuuden selvittämiseen. Tuote-
nimikkeiden kuukausi-, viikko-, viikonpäivä- ja kellonaikojen erojen tilastollinen merkit-
sevyys selvitettiin SPSS-ohjelmistolla. Ubersuggest-palvelusta saatiin tietoa tuotenimikkei-
den hakuja tehneiden kuluttajien ikäryhmistä, joka lisäsi kuluttajaymmärrystä. Boonen ym.
(2015) mukaan käytettäessä Internetiä ostopäätöksen tutkimiseen ajatellaan, että termien
hakeminen enteilee vastaavien tuotteiden tai tuotekategorioiden myyntiä.

Kehitysprosessin onnistuminen

Tutkimuksen tekijä vastasi tutkimus- ja kehittämisprojektin suunnittelusta, toteuttamisesta
ja raportoinnista. Tutkimus- ja kehitysprojektin ohjauksesta vastasivat kohdeyrityksen asi-
antuntijat ja opinnäytetyön ohjaaja. Projekti eteni projektisuunnitelman mukaan, johon oli
merkitty kriittiset pisteet projektin eri vaiheiden etenemisen kannalta. Pientä viivästystä ai-
heutti työn laajuus, joka muodostui arvioitua suuremmaksi. Ulkoisia ohjauspisteitä olivat
projektin käynnistyspalaveri, välitulosten esittely ja lopullisten tulosten esittely tapaami-
sissa. Eri vaiheiden välissä kysyttiin palautetta, joka otettiin huomioon jatkossa. Sisäisiä
ohjauspisteitä olivat opinnäytetyön ohjaustapaamiset ohjaajan kanssa.

Kehitysprosessin onnistumiselle oli tärkeää sopia heti projektin alkuvaiheessa tavoitteista, sisällöstä ja aikataulusta. Tutkimustyöstä oli vastuussa ONT-projekti, joka tuotti kehittämisprojektille tarvittavan tutkimustiedon. Kehittämisprojekti vastasi projektin hallinnasta. Ostopäätösprosessin sisäistäminen ja ymmärtäminen kohdeyrityksen liiketoiminnan kanalta oli oleellista tutkimuksen toteuttamiselle. Lähestymistapa auttoi pilkkomaan tutkimuksen selkeisiin vaiheisiin, jota edesauttoi kehitysprosessin etenemistä.

7.2 Tutkimuksen luotettavuus

Tutkimuksen luotettavuustarkastelun tarkoituksena on hallita tutkimusprosessin aikana syntyviä tutkijasta tai aineistosta johtuvia tietoisia tai tiedostamattomia virheitä, tehdä oikeita ratkaisuja ja perustella ne (Kananen & Makkonen (toim.) 2015, 338, 343). Luotettavuustarkastelulla pyritään aina mahdollisimman suureen luotettavuuteen. Luotettavuutta tarkastellaan tieteellisessä tutkimuksessa validiteetilla ja reliabiliteetilla (Kananen & Makkonen (toim.) 2015, 343.)

Validiteetti

Validiteetilla tarkoitetaan oikeiden asioiden tutkimista (Kananen & Makkonen (toim.) 2015, 343). Sitä voidaan pitää hyvänä, jos mittari onnistuu mittaamaan tarkoitettuja asioita (Taanila 31.3.2019). Tutkimuksen tulosten validiteetti on hyvä, sillä tutkittiin asioita, jotka oli määritelty tutkimuskysymyksissä. Peittomatriisiin (taulukko 1) on koottu tutkimusongelmiin liittyvät teoriat, tulokset ja aineistot, jotka osoittavat näiden välisen yhteyden. Tutkimuksessa huomioitiin koko prosessin luotettavuus. Liiketoiminnan ymmärtämisellä pyrittiin varmistamaan tutkimuksen validiteetti. Datatunneksen tarve määriteltiin liiketoiminnan vaatimuksesta lähtien, mutta saatavuus ja laatu heikensi validiteettia. Alkon kerätyn punaviinin myyntidatan validiteettia heikentää datasta puuttuvat matkustajatuonnin ja ulkomaisista verkkokaupoista tilattujen ostojen osuudet. Punaviinin osuus matkustajatuonnista on arviolta noin viisi miljoonaa litraa. Arvio perustuu Alkon (2020) eri viinilajien myyntitilastoihin ja Terveystieteiden ja hyvinvoinnin laitoksen (2019) tutkimukseen matkustajatuonnista. Sosiaali- ja terveystieteiden ministeriön asettaman työryhmän arvion mukaan verkkokaupan osuus alkoholin myynnistä on joitakin prosentteja (Sipinen 2018). Jos laskemme määrän 5 %:n oletuksen mukaan, on määrä noin yksi miljoonaa litraa. Yhteensä matkustajatuonti ja verkkokauppa-ostokset tekevät noin kuusi miljoonaa litraa punaviiniä, joka vastaa noin 21,6 prosenttia punaviinin vuotuisista ostoista Suomessa. Tämä tarkoittaa, että Google Trends -indekseissä on mukana myös sellaisia punaviinin hakuja, jotka eivät suoraan kohdistu Alkon punaviinin myyntiin. Puutteesta huolimatta Alkon punaviinin myyntidata sopi tutkimuksen mallinnukseen, koska tutkimuksen kohde ei ollut alkoholimäärän kulutus Suomessa.

Tutkimus- ja analyysimenetelmiä voidaan pitää valideina, sillä ne perustuvat teoriaan ja aikaisempiin tutkimuksiin. Datapisteiden määrä oli riittävän suuri analyysien toteuttamiseen. Lisäksi tutkimuksen ulkoinen validiteetti eli yleistettävyyys varmistettiin tutkimuksen eri vaiheissa testaamalla tulosten tilastollinen merkitsevyys. Tutkimustulokset antavat konkreettista tietoa kuluttajan ostopäätösprosessin tiedon etsimisen ja ostopäätöksen ajoittumisesta sekä myyntimäärän ennustamisesta. Yksityiskohtaisesti kuvattujen mallien avulla tuloksia voi soveltaa yrityksen liiketoiminnan uusiksi toimintamalleiksi.

Reliabiliteetti

Reliabiliteetilla tarkoitetaan samojen tulosten saamista tehtäessä tutkimus uudelleen (Kananen & Makkonen (toim.) 2015, s.343). Toisaalta mittauksen satunnaiset virheet huonontavat reliabiliteettia (Taanila 31.3.2019). Tämän tutkimuksen reliabiliteetti on hyvä, ja perustelut on esitetty seuraavassa prosessin eri vaiheissa. Tutkimuksen liiketoiminnan ymmärtämistä koskevassa osuudessa analysoitiin tietoa dokumenttien perusteella ja tehtiin puolistrukturoitu haastattelu. Tutkimuksen teoriaosuuden aineistot on kerätty monipuolisesti ja lähteitä kriittisesti tarkastellen. Aiheen ajankohtaisuuden takia on pyritty käyttämään mahdollisimman uutta lähdemateriaalia. Tietyt teoreettiset osuudet ja tilastolliset menetelmät on koottu myös vanhemmista, mutta yleisesti käytetyistä, luotettaviksi arvoiduista lähteistä. Lähteitä tarkasteltaessa huomioitiin myös mahdolliset sidokset, joten erilaisia asiantuntijänäkökantoja käytettiin monipuolisesti. Liiketoiminnan ja ilmiön ymmärtämiseksi taustatietoja kerättiin myös haastattelemalla. Haastattelutilanteessa vältettiin johdattelemista ja vastaukset nauhoitettiin kirjaamisvirheiden välttämiseksi.

Tutkimuksessa käytetyt datat ovat saatavana avoimesti Internetistä. Datat kerättiin huolellisesti ja varmistettiin niiden oikeellisuus keräämisen ja valmistelun eri vaiheissa. Alkon keräämä ja julkaisema alkoholijuomien tilastodata myynneistä on ensimmäisen osapuolen dataa. Data vaikuttaa erittäin luotettavalta, sillä alkoholilain (Finlex 2020) mukaan sosiaali- ja terveysalan lupa- ja valvontavirasto tuottaa tilastointipalvelut.

Google Trends -data on harhaton otos Google-hauista, joista vain prosenttia käytetään datassa (Google 2020b). Koska datan määrä on suuri, eri aikoina tehtyjen hakujen tulokset voivat hieman vaihdella keskenään (Stephens-Davidowitz & Varian 2015). Vaihtelu tuloksissa eri päivinä tehtyjen hakujen välillä on muutamia prosentteja (Choi & Varian 2011). Lisäksi Google Trends -datan arvo on pyöristetty kokonaisluvuksi. Jos tarvitaan tätä tarkempi lukuarvo tai muuten hyvin tarkkaa dataa, kannattaa tehdä useita Google Trends -hakuja ja laskea arvoista keskiarvo. Tämä tulee tehdä eri päivinä, sillä muutoin data tulee samasta otoksesta. Google Trendsin aikasarja- ja maantieteellinen data on normalisoitua, mikä vaikuttaa oikeellisuuteen. Tämä tarkoittaa, että jokainen numero on jaettu kaikkien

hakujen määrällä ja annettu korkeimmalle arvolle numeroksi 100. Jos alue A saa indeksin 100 ja alue B indeksin 50 samassa haussa nähdään, että hakutermiin sisältyvät hakujen prosenttiosuudet ovat kaksi kertaa suuremmat alueella A verrattuna alueeseen B. (Stephens-Davidowitz & Varian 2015.) Jos indeksi on 0, ei alueesta ole saatavana riittävä määrä hakutietoja (Google Trends 2020b). Google-dataa käytettäessä tulee huomioida, että äkilliset muutokset, päiväkohtaiset mallit ja tiettyjen alueiden suhteelliset muutokset ajan mittaan ovat merkityksellisempiä kuin pitkän ajan trendit, koska näihin voi liittyä hakukäyttäytymiseen liittyvä tulkintavirhe. (Stephens-Davidowitz & Varian 2015.)

Arvioitiin Ubersuggestin datan tarkkuutta ja luotettavuutta peilaamalla sitä kilpailevaan, maksulliseen hakukoneanalytiikkatyökaluun, SimilarWebiin, josta on tutkittua tietoa. Ubersuggestia voi käyttää kilpailijatiedon hallinnan työkaluna arvioitaessa eri verkkosivujen trafiikkia ja rankkauksia samaan tapaan kuin SimilarWeb -työkalua. SimilarWeb ilmaisee rankkauksen ja trafiikin absoluuttisina arvoina. Kun verrattiin kuukausittaisia verkkosivujen trafiikkeja, SimilarWeb erosi NetMonitorin arvoista keskimäärin 42 %. NetMonitoria pidetään tarkkana ja korkealaatuisena nettisivujen trafiikin mittarina. Esimerkiksi, jos todellinen trafiikki oli 100 000 vierailijaa kuukaudessa, SimilarWeb näytti määräksi 58 000–142 000 vierailijaa. SimilarWeb toimi tyydyttävästi (poikkeama enintään 30 % NetMonitoriin verrattuna) nettisivujen trafiikin arvioinnissa 49 %:ssa testatuista sivuista (n = 485). Mitä laajemmat nettisivut, sitä suuremmat olivat ongelmat trafiikin määrittämisessä: kun pienillä verkkosivuilla työkalu selviytyi tyydyttävästi 58 %:ssa sivuista laajoilla sivuilla arvo oli enää 38 %. Sekä NetMonitor että Google Analytics käyttävät sivustokeskeistä menetelmää verkkosivuston trafiikin keräämisessä, mutta taustalla oleva tekniikka eroaa toisistaan. Tulokset ovat kuitenkin yhtä tarkkoja. (Prantl & Prantl 2018.) Bhatin mukaan mikään avainsana-analyysityökalu ei ole tarkka, mutta Ubersuggestin tarjoama data on tarkkaa tiettyyn pisteeseen asti. Se tarjoaa melko luotettavaa dataa, koska datan lähde on Google. Ja jos datalähteet ovat luotettavat, voidaan myös tuloksiin luottaa. Tarkkaa reaaliaikaista dataa ei kuitenkaan ole saatavana. (Bhat 29.2.2020.)

Tutkimuksessa käytetyt tutkimusmenetelmät ja analyysit voi toistaa Excel-taulukkolaskentaohjelmalla ja SPSS-tilasto-ohjelmalla kuvausten perusteella, joten reliabiliteetti on hyvä. Sen sijaan reliabiliteettia heikentää muutaman prosentin ero tuloksiin, joka aiheutuu keräämisestä Google Trends- ja Ubersuggest-analytiikkatyökaluilla eri päivinä.

Tulosten arviointi perustuu erilaisiin lähteisiin, ja omat arvioinnit ja mielipiteet on eroteltu selkeästi.

7.3 Eettiset näkökohdat

Tutkimuksen laatimisessa on noudatettu Haaga-Helia ammattikorkeakoulun opinnäytetyön laatimisen eettisiä ohjeita (Haaga-Helia 2019c).

7.4 Johtopäätökset sekä kehittämis- ja jatkotutkimusehdotukset

Seuraavassa esitetään johtopäätökset kaikkiin tutkimusongelmasta johdettuihin kysymysmuotoisiin alaongelmiin (TK1–TK4) ja tehdään kehitys- ja jatkotutkimusehdotukset.

Tutkimuksen pääongelma on:

Milloin kuluttajat hakevat tietoa netistä päivittäiselintarvikkeista ja tekevät ostopäätökset, ja mitä dataa data-analytiikalla voi kerätä ja miten hyödyntää?

Pääongelmasta johdetut alaongelmat on muotoiltu tutkimuskysymyksiksi seuraavasti:

TK1: Mitä kuluttajaymmärrystä lisäävää dataa on saatavana ostoprosessin tiedon etsimisen ja ostopäätöksen vaiheista, ja mitä dataa data-analytiikan työkaluilla voi kerätä (huomioiden tiedon yksityisyyteen liittyvän lainsäädännön (GDPR))?

Dokumenttiaineistoista selvitettiin, millaisia datalähteitä B2B-sektorilla toimiva yritys, jonka merkittävät asiakkaat toimivat päivittäiselintarvikkeiden parissa, voi hyödyntää kuluttajaymmärryksen parantamiseksi data-analytiikalla. Asiaa tarkasteltiin kuluttajan ostopäätösprosessin näkökulmasta keskittyen tiedon etsimis- ja ostopäätös vaiheisiin. Tulokset on koottu kohtaan 6.1.

Tiedon etsimisvaiheessa keskeiseen rooliin nousi hakukonedata. Merkittävin datan lähde on Google, joka on Suomen ylivoimaisesti suosituin hakukone. Avointa Google Trends -työkalua voi hyödyntää massadatan keräämisessä. Sponderin ja Khanin (2018, 26) mukaan hakudata on avain kuluttajaymmärrykseen, koska hakukoneissa kuluttajat etsivät haluamiaan asioita.

Ostopäätös vaiheeseen on haastavaa löytää yrityksen ulkopuolista dataa elintarvikkeiden myynneistä ja ostoista. Tilastokeskukselta on saatavana avointa dataa elintarvikkeiden kulutuksesta ja kuluttajien taustatiedoista. Tiedot vähittäiskaupan myyntitiedoista löytyvät kaupparyhmiltä, mutta ainakaan tällä hetkellä tietoja ei luovuteta eikä myydä ulkopuolisille yrityksille. Tilastokeskus suunnittelee ostaja- ja tuotedatan yhdistämistä kaupan ja pankin datoista tilastointia varten, ja vastaavaa mallia on jo testattu Norjassa. Datat kerääminen tilastointiin vaatii asiakkaan luvan, joten sen saaminen yrityskäyttöön lähiaikoina vaikuttaa

epätodennäköiseltä. Systemaattisesti kerättävää ja julkaistavaa avointa vuositason tarkempaa dataa elintarvikkeiden myynnistä oli saatavana ainoastaan Alkon alkoholijuomien myyntitilastoista. Tiedon saatavuus on tiedon siirtämisen rajapintoja suurempi haaste tällä hetkellä. Tietyillä maksullisilla ohjelmilla pääsee kiinni kuluttajadatoihin, ja maksullisella toisen ja kolmannen osapuolen datalla voi rikastaa omaa dataa. Paneelidatasta saa esimerkiksi demografiatietoja ja väestöestimaatit. Myyntidataan yhdistetty paneelidata lisää ymmärrystä asiakkaan mieltymyksistä ja käyttäytymisestä. Omat innovatiiviset digitaaliset palvelut ovat avain kuluttajätietoihin, sillä kuluttajat antavat luvan nettikäyttäytymisensä seurantaan ja tietojensa keräämiseen, jos kokevat saavansa siitä riittävästi hyötyä.

Tutkimuksessa nousi esille jatkotutkimusehdotuksena selvitys, millaista maksullista kohdeyritystä hyödyntävää dataa on saatavana joko yrityksiltä tai analytiikkatyökalujen avulla. Nämä kannattaa kartoittaa ja miettiä hyödyntämismahdollisuudet olemassa olevan datan pohjalta kuluttajaymmärryksen lisäämiseksi. Myös kuluttajätietojen keräämismielessä kehitettävä oma digitaalinen palvelu olisi mielenkiitoinen jatkokehitysprojekti.

TK2: Milloin kuluttajat hakevat tietoa Googlesta valituista päivittäiselintarvikkeista? Miten trendit, sesongit, kuukaudet, viikot, viikonpäivät, kellonajat ja alueet vaikuttavat eri päivittäiselintarvikkeiden Google-hakuihin? Miten kuluttajaryhmät eroavat tiedonhakemisessa Googlesta eri päivittäiselintarvikkeiden kohdalla?

Ostoprosessin tiedon etsimisvaiheeseen liittyvät päivittäiselintarvikkeiden hakuajankohtia koskevat tulokset on esitelty kohdassa 6.2., jossa on myös tiedot arvojen 95 %:n luottamusväleistä. Elintarvikeryhmät purettiin tuotenimikkeiksi ensisijaisesti COICOP-HBS-luokituksella, datat kerättiin Google Trends -työkalulla ja käsiteltiin Excel-taulukkolaskentaohjelmalla sekä luottamusvälit laskettiin SPSS-ohjelmalla. Lisäksi selvitettiin Ubersuggest-työkalulla kuluttajien käyttämät hakusanat ja -volyymit sekä eri-ikäisten kuluttajaryhmien erot tiedonhakemisessa.

Google Trends -hakuindekseistä saatiin ajankohdat, milloin kuluttaja etsivät tietoa Googlesta valituista päivittäiselintarvikkeista. Selvitetiin päivittäiselintarvikeryhmien tuotenimikkeiden hakemisen trendit, sesongit, kuukaudet, viikot, viikonpäivät, kellonajat ja hakualueet.

Tutkimuksessa havaittiin, että elintarvikeryhmistä kaikki tutkitut meijerituotteet sekä kasvi-proteiineista soijarouhe olivat trendinä nousevia. Sen sijaan valmisruoista maksalaatikko, kasviproteiineista Nyhtökaura ja Härkis sekä kaikki tutkitut joulusesonkituotteet (porkkana-

ja lanttulaatikko sekä joulukinkku) olivat trendinä laskevia. Kausiluonteisuus näkyi selkeästi sesonkituotteiksi nimetyn ryhmän kohdalla, jossa vuoden ajanjaksolla erottui selkeästi yksi hakupiikki, kun muutoin hakuja oli hyvin vähän tai ei ollenkaan. Joulua edeltävät viikot 50–51 näkyivät hakuhuippuna sesonkituotteiden ohella myös voin, juuston, maksalaatikon ja punaviinin hakemisessa.

Unileverin ja USA:n suurimman vähittäistavaraketju Krogerin mukaan maanantait ovat tärkeitä viikon ruokalistojen suunnittelupäiviä (Kotler & Keller 2016, 196). Suomessa sunnuntai voi olla maanantaita tärkeämpi seuraavan viikon ruoanvalmistuksen suunnittelupäivä, koska aterian osan muodostavia Nyhtökauraa ja Härkistä haettiin silloin eniten. Ei-sesonkituotteista alkuviikon arkipäivinä haettiin erityisesti maanantaina maksalaatikkaa ja tiistaina jauhelihamaisesti käytettävää soijarouhetta. Viikonloppuisin haut kohdistuivat enemmän herkutteluun, sillä silloin haettiin intensiivisimmin voita, jäätelöä, juustoa ja punaviiniä. Kaikkia ei-sesonkituotteita haettiin eniten iltapäivisin aikahaarukassa klo 15–17.

Kuluttajaryhmien välisistä eroista saatiin tietoa ainoastaan ikäluokittain. Ikäluokista 18–24-vuotiaat hakivat eniten jäätelöä, soijarouhetta ja punaviiniä, 25–34-vuotiaat lanttulaatikkaa ja joulukinkkua sekä 35–44-vuotiaat voita, juustoa, maksalaatikkaa, Nyhtökauraa, Härkistä ja porkkanalaatikkaa. Alle 18-vuotiailta ja yli 65-vuotiailta ei ollut ollenkaan hakuja tai hakujen määrä oli niin pieni, ettei se näkynyt tilastoissa. Ikäryhmistä 55–64-vuotiaat hakivat Googlestä enemmän tietoa päivittäiselintarvikkeista kuin 45–54-vuotiaat, vaikka ikäryhmä ei noussut kärkeen minkään tuotenimikkeen hakemisessa.

Alueellisesti hakuintensiteetti vaihteli vuodesta ja maakunnasta toiseen kaikilla tuotenimikkeillä. Soijarouhetta haettiin viiden vuoden tutkimusjaksolla vain seitsemässä maakunnassa.

Voita haettiin eniten lauantaina, mutta ero oli tilastollisesti merkitsevä ainoastaan maanantaihin. Suosituin haku aika oli illansuussa klo 17, joka erosi tilastollisesti esimerkiksi tiettyihin aamun tunteihin. Jäätelöä haettiin useimmin touko-, kesä- ja heinäkuussa, joka erosi merkitsevästi alkuvuoden kuukausien kanssa. Jäätelön hakuhuippu osui viikonloppuun eli perjantaihin, lauantaihin ja sunnuntaihin, mutta viikonpäivien välillä ei ollut tilastollisesti merkitseviä eroja. Eniten jäätelöä haettiin iltapäivällä klo 16–19, mikä erosi merkitsevästi yöstä ja aamusta. Kesähelteiden kuumien aikojen, klo 16–18, osuu samalle aikajänteelle (Ilta-sanomat 2015). Juuston hakuintensiteetti on suurimmillaan lauantaina, joka eroaa merkitsevästi keskiviikosta sekä klo 16–18, joka eroaa merkitsevästi aamusta.

Valmisruokia haettiin Google Trendisistä niin sanotuilla long tail -hakusanoilla, jotta saataisiin yhdistettyä valmisruokanimikkeeseen valmistaja ja karsittua reseptien hakijat pois hakutuloksista. Ubersuggest-palvelussa varmistettiin, että kaikista yhdistelmistä löytyy ainakin jonkin verran hakuja. Kinkkukiusauksesta ja kasvisseosekeitosta ei saatu indeksien arvoja, sillä hakujen määrä ei riittänyt kyseisillä hakutermeillä. Muutama kymmenen hakua ei siis riitä indeksin arvoon.

Havaittiin, että maksalaatikko kuuluu monen joulupöytään, sillä sitä haettiin eniten joulukuussa viikolla 50. Maksalaatikko oli useimmin mielessä maanantaisin hakujen perusteella, mutta keskiviikkoisin ei hakuja ollut juuri ollenkaan. Eri viikonpäivien välillä ei ollut kuitenkaan tilastollisesti merkitsevää eroa. Eniten hakuja osui keskimäärin iltapäivään klo 16, mikä erosi merkitsevästi iltayhdeksästä.

Nyhtökauraa haettiin eniten maaliskuussa, mutta merkitsevät erot hakemisessa osuivat muihin kuukausiin. Suosituin hakupäivä oli sunnuntai, mutta eri viikonpäivien välillä ei ollut merkitseviä eroja. Hakupiikki osui iltapäivään klo 17, joka erosi merkitsevästi myöhäisillasta, yötä ja varhaisesta aamusta. Soijarouheen suurin hakuintensiteetti osui tammi-kuulle, mikä erosi merkitsevästi esimerkiksi kevätkuukausista. Viikonpäivistä sitä haettiin eniten tiistaina, mutta merkitseviä eroja eri viikonpäivien välillä ei ollut havaittavissa. Suosituin haku-aika oli klo 15 iltapäivällä, mikä erosi merkitsevästi tietyistä yön, aamun ja illan tunteista. Härkiksen suurin hakupiikki osui syyskuuhun. Eri kuukausien välillä oli merkitseviä eroja. Viikonpäivistä suosituin hakupäivä oli sunnuntai, mutta tilastollisia eroja muihin päiviin ei ollut havaittavissa. Eniten Härkistä haettiin iltapäivällä klo 16–17, joka erosi merkitsevästi esimerkiksi yön ja varhaisaamun tunteista.

Kaikkia joulusesonkituotteita haettiin eniten sunnuntaina ja maanantaina eli 1–2 päivää ennen jouluaattoä, joka vuonna 2019 oli tiistai. Tulos on samansuuntainen Lidlin myyntitietojen kanssa. Lidlin mukaan jouluostokset tehdään pääsääntöisesti 22.–23.12. eli 1–2 päivää ennen jouluaattoä. Jouluaaton viikonpäivä ei vaikuta ostosten tekemisen ajankohtaan. (Tuominen 2019.) Hakemisessa ei ollut kuitenkaan tilastollisesti merkitsevää eroa eri viikonpäivien välillä.

Verrattaessa joulukinkun Google Trends -hakuindeksejä vähittäiskaupan myyntitietoihin huomataan, että suurin kinkun myyntihuippu osuu 19.–21.12. eli 3–5 päivää ennen jouluaattoä (Yhteishyvä 2020), kun taas joulukinkkua haetaan eniten varhain jouluaattoamuna klo 5, mutta aika ei eroa merkitsevästi muista kellonajoista. Ajankohta viittaa siihen, että joulukinkun hakemista ei tehdä aattona ostamisen takia, vaan Googlasta etsitään tuolloin

valmistus- ja/tai paisto-ohjetta. Joulukinkun ostaminen vaatii ennakointia, sillä pakastekinkun sulattamien kestää 2–5 päivää kinkun painon mukaan. Tuoreen kinkun suolaukseen varattava aika on puolestaan 7–10 päivää. Tuoresuolatun kinkun osto- ja paistopäivä kannattaa osuttaa lähelle toisiaan niin, että kinkku ostetaan enintään 2–3 päivää ennen paistamista. (Valio 2019.)

Sekä porkkana- että lanttulaatikka haettiin eniten sunnuntaina 22.12. eli kaksi päivää ennen jouluaattoa. Porkkanalaatikon suurin hakuintensiteetti osui aamupäivään klo 10–11, ja oli suurta myös klo 12–14 ja 16. Hakuhuippu erosi merkitsevästi yöstä ja aamuyöstä. Lanttulaatikka haettiin eniten klo 12, mutta erittäin paljon myös klo 9–11, 13–15 ja 17–18. Aamun ja illan hakuajoissa on merkitsevää eroa yöhön. Laatikoiden hakuajankohdat viittaavat vahvasti ostoaikeisiin, sillä Lidlin mukaan joulun ostokset tehdään pääsääntöisesti 22.–23.12. klo 12–16 (Tuominen 2019).

Punaviini oli trendinä nousujohteinen. Sitä haettiin ympäri vuoden, mutta selkeästi eniten ennen joulua viikoilla 50 ja 51. Joulukuu erosi merkitsevästi muista kuukausista hakemisessa. Punaviiniä haettiin eniten lauantaisin, joka erosi merkitsevästi keskiviikosta. Vaikka punaviini ei ole pelkästään kausiluonteinen tuote, katsottiin sen hakuintensiteetit joulun aikaan, koska Alkon mukaan jo useina vuosina eniten myynyt tuoteryhmä jouluna on ollut punaviini. Alko ennustaa myyvänsä ennen joulua kahden viikon aikana punaviiniä yli 2 miljoonaa litraa. (Baraka 2019.)

Punaviinin hakuindeksejä tarkasteltiin suhteessa Alkon aukioloaikoihin (Valitut Matkat 2019) joulun aikaan. Punaviiniä haettiin perjantaina 20.12. erityisesti klo 17–21, jolloin Alko oli avoinna klo 21 asti. Seuraavana päivänä lauantaina 21.12. punaviiniä haettiin eniten klo 14 ja 16–18 Alkon ovien mennessä kiinni klo 18. Sunnuntaina 22.12. Alko oli suljettuna, joka näkyi myös notkahduksena hakuintensiteetissä. Aatonaattona maanantaina 23.12. punaviiniä haettiin eniten klo 12–18. Tällöin Alko oli avoinna klo 21 asti. Jouluaattona 24.12. Alko piti ovensa auki klo 9–12, mutta joulunajan suurin hakuintensiteetti osui sulkemisajan jälkeiseen aikaan klo 14–17. Joulupäivänä 25.12. Alko oli suljettuna, mutta klo 14 punaviiniä haettiin taas suurella kiinnostuksella. Jouluaattoiltapäivän ja joulupäivän haut voivat arveluni mukaan liittyä joululahjaksi saatujen tuotteiden tutkimiseen. Vähittäiskaupat voivat myydä alkoholijuomia (2,8–5,5 %) klo 9–21.

Kohdeyritys voi hyödyntää Google Trendsistä kerättyjä päivittäiselintarvikkeiden hakuajan-kohtia koskevia tutkimustuloksia monella tavalla kuluttajan ostoprosessin tiedon etsimisen vaiheessa. Tässä vaiheessa voidaan vahvistaa brändiä ulkomainonnan avulla. Brändin

tunnettuus tukee kuluttajan ostopäätöstä, kun hän päättää ostaa juuri tietyn tuotteen. Ensiksi yritys voi tunnistaa trendit ja sesonkiluonteisuuden, ja hyödyntää niitä pitkän aikavälin suunnittelussa. Kun tiedetään milloin eri elintarvikeryhmät ja tuotteet kiinnostavat kuluttajia, voidaan ajankohdat merkitä vuosikelloon ja suunnitella myyjien kontaktointi mainostaviin asiakkaisiin optimaalisella hetkellä esimerkiksi kaksi kuukautta ennen ulostuloa. Toisaalta hakuintensiteetistä nähdään, milloin tuotteet eivät kiinnosta kuluttajia, jolloin kampanjoilla voi tasata kysyntää. Lisäksi pitkän aikavälin Google Trends -hauista voidaan seurata uusia nousevia trendejä ja löytää uusia asiakkaita. Suunnitelmat voidaan tehdä tulosten perusteella sesonkituotteissa jopa päivämäärän tarkkuudella.

Google Trendsin viikkodatasta saadaan selville lyhyen aikavälin haut, jolloin nähdään minä viikonpäivänä ja kellonaikana kuluttajat ajattelevat tiettyjä tuotteita. Tiedon haku Internetistä enteilee ostopäätöstä. Googlen hakudatan ja oston välisen viiveen avulla voidaan arvioida ostopäätöksen ajankohta. Arviointi voi perustua myös tutkimuksiin tai tietoon myyntiajankohdasta. Kun kohdeyritys esittää perustellun arvion optimaalisesta mainostusajankohdasta, se voi tukea asiakkaansa päätöksentekoa. Onnistuminen voidaan todentaa mittaamalla. Kun kuluttaja näkee ulkomainontaa tuotteesta ostopolkunsa varrella lähellä ostopäätöshetkeä, se aktivoi häntä ostamaan tietyn tuotteen. Lyhyen aikavälin Google Trends -datasta voi seurata myös ulkomainonnan tehokkuutta ohjata Googleen hakemaan tietoa mainostetusta tuotteesta. Jos seurantaa tehdään enintään edellisen päivän datasta päästään minuuttitarkkuuteen ja enintään viikkoa vanhasta datasta päästään tuntien tarkkuuteen. Jos ulkomainonta ohjaa suoraan yrityksen nettisivuille tai kampanjasivuille hakemaan lisätietoja tuotteesta, voidaan asiakkaan sivujen trafiikkia seurata esimerkiksi Ubersuggest-työkalulla. Kohdeyrityksen kannattaa hyödyntää Ubersuggest-hakuja myös täydentämään Google Trends -hakutietoja hakuvolyymeilla sekä hakuja tehneiden henkilöiden ikäryhmillä ja päätelaitetiedoilla. Lisäksi palvelussa voi täsmentää hakutermit, jotta tiedetään, millä hakusanoilla ihmiset oikeasti hakevat asioita.

Google Trends -datan rajoite kohdeyrityksen kannalta on sen aluekohtainen tarkkuus, joka rajoittuu enintään kaupunkitasolle. Google Trends -dataa ei suoraan voi kohdistaa tarkkaan paikkaan korttelitasolla. Toisaalta kuluttajien hakukäyttäytyminen eri alueilla, esimerkiksi kaupungin sisällä, voi olla riittävän samansuuntaista, jotta datan käyttö on hyödyllistä.

Tutkimuksessa nousi esille muutama jatkotutkimusehdotus. Kohdeyrityksen olemassa olevan paikkatiedon yhdistäminen Google-hakuajankohtiin olisi kiinnostava jatkotutkimus. Tällöin voitaisiin arvioida, riittääkö Google Trendistä -saatavan alueita koskevan hakutie-

don tarkkuus kohdeyrityksen käyttöön. Lisäksi voisi innovoida, millaisen verkko- tai mobiilipalvelun avulla kuluttajat suostuisivat antamaan luvan tietojen keräämiseen ja hyödyntämiseen. Tämä olisi erityisen hyödyllistä, jos käytettävissä olisi myös kuluttajan ostopolun varrella olevien liikkeiden maksupäätedata, josta saisi selville tuotetiedot ja pankkien transaktiodata, josta saisi tiedot ostajista.

TK3: Kuinka pitkä viive on punaviinin Google-hakujen ja ostamisen välillä Suomessa ja eri maakunnissa?

Kahden aikasarjamuuttujan, Google Trends -indeksidatan ja punaviinin myyntidatan, risti-korrelaatioanalyysillä tutkittiin ostoprosessissa punaviinin Google-haun ja ostamisen välistä viivettä, jolla päästiin kiinni ostopäätöksen ajankohtaan. Ostamista tutkittiin myyntitiedoilla, joiden ajankohta olisi sama kuin ostotiedoilla, jos sellaiset olisi saatavana. Koska Alkon myyntidataa oli saatavana vain kuukausitasolla, myös tulokset ovat tällä tarkkuudella. Tulokset ovat nähtävissä kohdassa 6.4. Mahadevan (2010, teoksessa Kinski 2016, 16) mukaan lyhyen aikavälin ennustuksen aikajänne on 1–3 kuukautta, mutta Choin & Varianin (2012) mukaan lyhyen aikavälin ennustaminen tarkoittaa nykyisyyden ennustamista, joka heidän esittämässään ennustemallissa ylittää kolmen viikon päähen. Kuukausidatalla voitiin siis tehdä lyhyen aikavälin ennuste, mutta nykyisyyden ennustamiseen olisi tarvittu vähintään päivadataa. Suomessa havaittiin voimakas positiivinen korrelaatio 95 % luottamusvälillä punaviinin Google Trends -hakuindeksien ja myyntien välillä saman kuukauden aikana. Voimakas positiivinen korrelaatio punaviinin hakuindeksin ja myynnin välillä viittaa siihen, että molemmat muuttujat reagoivat samalla tavalla ja samaan aikaan johonkin ulkoiseen tekijään (IBM 2020a). Muuttujien välillä oli vielä kohtalainen positiivinen korrelaatio seuraavan ja sitä seuraavan kuukauden aikana 95 % luottamusvälillä. Tuloksia tukee tutkimustieto, että Pohjoismaissa 86 % kuluttajista (n = 30 000) tekee päivittäistavarojen ostopäätöksen enintään 5 kuukautta aikaisemmin. Tästä joukosta 44 % tekee ostopäätöksen aikaisintaan edellisellä viikolla. (Kuulas Helsinki 2019.) Tutkimuksessa kahden elintarvikkeen, tapaksen ja paellan, Google Trends -hakuintensiteetin ja ostojen välillä ei havaittu merkitsevää viivettä. Mallin viikkomyynnteihin vaikuttivat voimakkaasti saman viikon hinta ja edellisen viikon myynnit, mutta kahta viikkoa suurempi viive vaikutti vähemmän tuotteiden myyntiin. (Boonen ym. 2015.) Tulokset osoittavat, että tutkimuksen kuukausidata ei sovellu viiveen ennustamiseen, kun ennustetaan nykyisyyttä enintään kolmen viikon aikajänteellä. Data oli kuitenkin käyttökelpoista mallinnukseen, jolloin mallia voi hyödyntää sellaisenaan, jos optimaalista päivä- tai tuntidataa on saatavana.

Maakuntien välisiä eroja tutkittiin spatiotemporaalisella analyysillä kuvaamalla punaviinin Google Trends -hakuindeksien ja myynnin välisiä viive-eroja kartoilla. Tilastollisesti merkitsevä kohtalainen positiivinen lineaarinen riippuvuus havaittiin seitsemässä maakunnassa saman kuukauden aikana. Seuraavan kuukauden aikana merkitsevä positiivinen lineaarinen riippuvuus oli havaittavissa enää kuudessa maakunnassa ja sitä seuraavan kuukauden aikana enää Uudellamaalla. Tutkimuksessa osoitettiin, että Google Trends dataa voi käyttää yhdessä myyntidatan kanssa ennustamaan eri alueiden välisiä viive-eroja, jotka voi visualisoida kartoiksi. Viive-erot ennustavat ostopäätösajankohdan hakuindeksien ja myynnin historiadata perusteella.

Kohdeyritys voi hyödyntää mallia ostopäätösajankohdan selvittämiseksi, jos saa käsiinsä esimerkiksi kumppaneiden kautta riittävän tarkkaa myyntidataa, joka käytännössä tarkoittaa tuntidataa. Tiedon avulla mainostava asiakas voi aktivoida ulkomainonnalla kuluttajaa tekemään tuotteensa ostopäätöksen ostopaikalla tai sen lähellä. Nykyisyyden ennustaminen enintään kolmen viikon päähän, olisi myös riittävän pitkä aika mainostavan asiakkaan kannalta valmistella päivittäiselintarvikemainos. Viiveen pituus tiedon etsimisestä ostopäätökseen vaihtelee eri tuotteilla, kun auton ostoa mietitään yleensä kuukausia, tehdään päivittäistavaroita koskeva päätös yleisimmin enintään edellisellä viikolla. Kohdeyritykselle kertyvä historiadata auttaa ennusteen tarkkuuden parantamisessa. Kohdeyritys voi visualisoida eri alueiden väliset viive-erot spatiotemporaalisten karttojen avulla. Kartoilla näkee alueiden väliset erot yhdellä silmäyksellä. Kohdeyritys tarvitsisi kuitenkin maakuntia tarkempaa alueellista tietoa myynnin tueksi.

Tutkimuksessa nousi esiin muutamia jatkotutkimusehdotuksia. Mallia kannattaa testata vähintään tuntidatalla, jotta nähdään sen toimivuus ilmiön kannalta optimaalisissa olosuhteissa. Lisäksi kohdeyrityksen kannattaa tutkia suosituimpien mainostettavien tuotteiden viiveet, jotta saadaan selville näiden todennäköiset ostopäätösajankohdat. Näin voitaisiin päätellä Google Trendsin hakuintensiteetin perusteella tulevat ostoajankohdat eri tuotetyypeillä. Lisäksi olisi hyödyllistä tietää, miten ostopäätösajankohdan voisi yhdistää paikkatietoon.

TK4: Onko punaviinin Google-hakuintensiteetillä ja myynnillä positiivinen riippuvuus Suomessa, ja voiko niiden perusteella ennustaa myyntimäärää?

Aluksi selvitettiin muuttujien, Google Trends -indeksien ja punaviinin myynnin, välinen riippuvuus, jotta voitiin valita analysointimenetelmät. Menetelmäksi valittiin Pearsonin korrelaatioanalyysi. Havaittiin, että Google Trends -indeksien ja punaviinin myynnin välillä on voimakas positiivinen riippuvuus Suomessa tilastollisesti merkitsevästi. Koska muuttujien

välillä oli riippuvuus, voitiin käyttää lineaarista regressiomallia myyntimäärien ennustamiseen, jolloin voitiin hyödyntää aikasarjojen toteutuneita myyntejä ennustamisessa. Tulokset on esitetty kohdassa 6.4.

Lineaarisen regressiomallin ennustemäärät punaviinin myynnille laskettiin Google Trends -indeksin arvoille, jotka mahtuivat havaintoalueelle. Verrattaessa ennusteen indeksien ääriarvoja historiatietoihin vaikuttaa siltä, että ennusteen arvot ovat liian suuria indeksin skaalan yläosassa, vaikka myynnin nouseva trendi otettaisiin huomioon. Tämä johtuu arvioni mukaan siitä, että regressiomallia ei ole korjattu kausivaihtelulla. Punaviiniä myydään ympäri vuoden, mutta myynnissä näkyy selkeä kausivaihtelu ennen joulua, jolloin myyntimäärä lähes kaksikertaistuu keskiarvon kuukausimyyntiin nähden. Punaviinin Google-hakujen huippu osuu myös joulukuuhun. Keskiarvon luottamusväli on erittäin laaja, etenkin yläarvolla yli kaksi kertaa keskiarvon verran 95 %:n varmuudella. Tosin tämä kuuluu jossakin määrin asiaan, sillä ennusteet käyvät sitä epätarkemmiksi, mitä kauemmaksi keskiarvoista mennään (Taanila 2010, 17). Mallin hajontakuviosta nähtiin, että indeksin arvolla > 30 esiintyi runsaasti muista poikkeavia havaintoja ennustetun myynnin yläarvoilla. Poikkeamat eivät johtuneet virheellisistä tai väärin syötetyistä tiedoista, joten niitä tulee tarkastella tutkittavan ilmiön näkökulmasta. Kuten edellisessä kohdassa AO3 todettiin, päivittäiselintarvikkeiden viive hausta ostoon on melko lyhyt. Google Trends -indeksien arvot keskittyivät välille 10–40, mutta suuremmilla arvoilla oli yksittäisiä datapisteitä. Tulosta olisi voinut parantaa poistamalla yksittäisiä arvoja, mutta lähtökohtaisesti myyntidata olisi pitänyt olla kuukausidataa tarkempaa.

Koko malli sopi aineistoon tarkasteltaessa selitetyn riippuvan muuttujan vaihtelun ja jäännöstermin suhdetta, koska koko malli oli tilastollisesti merkitsevä. Myynnin vaihtelusta pystyttiin selittämään kuitenkin vain vajaat puolet hakuintensiteetin vaihtelulla. Mallin selitysaste ei kuitenkaan ole ainut tekijä, joka kuvaa mallin toimivuutta käytännössä. Myös peruste, jolla selittävä muuttuja on valittu mukaan malliin, kuvaa koko ilmiön ymmärtämistä. Google Trends -indeksin painoarvo mallille nähdään selittävän muuttujan normitetusta beta-kertoimesta. Koska se oli positiivinen, kasvattaa hakuintensiteetti punaviinin myyntiä. (Nummenmaa 2011, 328.) Mikäli mallissa olisi mukana useita selittäviä muuttujia, voisi beta-kertoimien arvosta verrata muuttujien selitysarvoa myyntimäärille (Taanila 2010, 10). Ylipäänsä selittävä muuttuja sopi malliin, ja se oli merkitsevä selittäjä. Aikaisemmin toteutetun tutkimuksen mukaan, Google Trends -datalla täydennetty yksikertainenkin regressiomalli parantaa taloudellisten muuttujien ennustekykyä 20 % verrattuna päivittäiskaupan ennustemalleihin, joissa ei käytetä Google Trends -dataa (Choi & Varian 2009). Jäännös- virheellä tarkasteltiin punaviinin myynnin osaa, jota mallilla ei voitu ennustaa eli mallin en-

nustaman ja toteutuneen havaintoarvon erotusta. Virhearvion vaihtelu keskiarvoon nähdessä oli melko suurta etenkin skaalan yläpäässä. Samoin jäännöstermien itseisarvot olivat aika suuria suhteessa punaviinin myynnin keskiarvoon, joten malli ei pysty ennustamaan myyntiä kovin tarkasti. Ennustemallin keskihajonta oli kuitenkin pienempi kuin historiatietoihin perustuvilla myynneillä.

Kokeiltiin vielä parantaa mallin selitysasetta epälineaarilla käyräsovituksella. Mallin selityssaste parani kuutiomallilla yli kuusi prosenttia. Kuutiomalli huomioi paremmin suurimpien indeksien vaikutuksen punaviinin myyntiin, koska pisteet osuvat tällöin lähemmäksi käyrää. Verrattaessa koko mallin tilastollista merkitsevyyttä F-arvolla havaittiin, että lineaarinen malli oli kuitenkin kuutiomallia parempi, koska kuutiomallissa on kolmannen asteen yhtälönä enemmän parametrejä, jotka tekevät mallista monimutkaisemman. (Mauranen 2020.)

Regressiomallin tilastollinen merkitsevyys testattiin edeltävyysehdoilla. Aluksi tarkasteltiin hajontakuviolla lineaarisuutta ja jäännösten variansseja. Mallin jäännöskuviossa pisteet jakautuivat x-akselin molemmin puolin melko tasaisesti, joten jäännöstermit viittaavat mallin toimivuuteen ja normaaliin jakaumaan. Toisaalta pisteet painottuivat voimakkaasti y-akselin suuntaisesti noin yhden arvon päähän nolasta kumpaankin suuntaan. Pisteet eivät jakaudu säännönmukaisesti jäännöskuviossa, joten kuvio viittaa lineaarisuuteen. (Nummenmaa 2011, 324–325.) Poikkeavia havaintoja esiintyy y-akselin suuntaisesti kaukana nolasta. Pistejoukossa ei näy kiilamaista kasvua, jota tapahtuisi, jos jäännöstermien varianssi kasvaisi ennusteiden kasvaessa. (Taanila 2010, 13–14.) Arvelen, että havaintoarvojen keskittyminen nolalla lähelle johtuu pitkälle aineiston luonteesta, sillä esimerkiksi Boonen ym. (2015) tutkimuksen mukaan ei löydetty merkitsevää viivettä tutkittujen elintarvikkeiden Google Trends -hakuintensiteetin ja ostojen välillä. Koska Google-haku ja osto tapahtuvat lähellä toisiaan, ei Alkon punaviinin myyntidata ole lähtökohtaisesti optimaalinen tutkimukseen, koska se on kuukausidataa. Ostamista indikoidaan tässä myyntitietojen avulla. Vaikka malli ei puutteistaan huolimatta ole paras mahdollinen, ei se kuitenkaan riko suoranaisesti lineaarisuutta tai jäännösten varianssien yhtäsuuruutta. Regressiomalli on aina loppupeleissä kompromissi selittävien muuttujien kriteerien, käytännön toimivuuden ja edeltävyysehtojen välillä (Taanila 2010, 18).

Seuraavaksi tarkasteltiin jäännösten normaalijakautuneisuutta vertaamalla havaintoja suoraan. Kuviossa havaittiin jonkin verran poikkeamia suoraan verrattuna, mutta ei vakavia eroja. (Taanila 2010, 14–15.) Tulos tarkistettiin vielä vinous- ja huipukkuuslukujen avulla. Myyntijakauman vinous ja huipukkuus olivat tilastollisesti merkitseviä, joten jäännökset ei-

vät ole tämän perusteella normaalijakautuneita. (Saaranen 2018, 18–19.) Toisaalta otoskoko oli riittävän suuri (> 30) normaalijakaumaoletuksen kannalta (Taanila 2019a, 5), mutta muuttujan jakauman ollessa epätavallinen, esimerkiksi erittäin vino, tulee otoskokoa kasvattaa normaalijakautuneisuuden toteutumiseksi (Taanila 2019a, 12). Jakauman normalius ei ollut aivan yksiselitteinen ja selkeä, mutta pienet poikkeamat eivät pääsääntöisesti ole vakavia. Isoilla poikkeamilla merkitsevyystestit ja ennusteiden luottamusvälit menettävät pätevyytensä. Tällaisessa tapauksessa mallia tulee korjata muuttujien muunnoksilla. (Taanila 2010, 15.) Toisaalta otoskoon ylittäessä 30, normaalijakautuneisuuden testausta ei tarvitse tehdä (Taanila 2019a, 12), joten johtopäätökseni on, että poikkeamat normaalijakaumassa eivät olleet vakavia, eikä tilastollinen merkitsevyys vaarannu. Google Trends -indeksien ja punaviinin myynnin välillä ei ollut autokorrelaatiota eli jäännökset olivat toisistaan riippumattomia.

Lopuksi voisi todeta, että malli ei tällaisenaan sovellu tarkkaan punaviinin myynnin ennustamiseen, mutta se on hyvä lähtökohta aloittaa mallin muokkaaminen paremmaksi. Aluksi myyntidatan tulee olla tarkempaa eli vähintään päivä-, mutta mieluiten tuntidataa. Regressiomallia voi parantaa esimerkiksi huomioimalla kausivaihtelu. Aikaisemmassa tutkimuksessa kausivaihtelun huomioivalla autoregressiivisellä mallilla osoitettiin Google Trends -datalla, että yhden prosentin kasvu automerkin hakuintensiteetissä kasvatti myyntiä noin 0,5 prosenttia (Choi & Varian 2009). Regressiomallia parantaa myös tuotteen myyntiin oleellisesti vaikuttavien selittävien muuttujien lisääminen. Ulkomainonnassa malliin tulee rakentaa mukaan oikea yleisö, paikkatieto, ajankohta ja altistuminen mainonnalle, joka on yhteismitallinen mainonnan mittari.

Toimiva ennustemalli parantaa kohdeyrityksen kilpailukykyä, koska se voi mallin avulla osoittaa mainostavalle asiakasyritykselle millaisia myyntimääriä on odotettavissa, kun tuotteen Google-hakuintensiteetti on tietyllä tasolla. Myös ostopolulle sijoittuvat kaupat hyötyvät mallista, kun liikkeet osaavat varautua oikealla määrällä tuotteita. Näin liikevaihto saadaan maksimoitua, kun tavara ei lopu kesken ja samalla voidaan optimoida myyntikate, kun ylimääräisiä tavaroita ei tarvitse myydä alennuksella. Ennustemalli lisää myös kuluttajatytyväisyyttä, kun tuote on saatavana kuluttajan haluamana ajankohtana.

Tutkimuksessa nousi esiin, että jatkotutkimuksella kannattaa selvittää tarkemman, päivätai tuntidatan vaikutus mallin ennustekykyyneen. Samoin on hyvä testata mallin korjaamisen vaikutus kausivaihtelulla, jos tuote reagoi kausivaihteluun, ja myyntiin olennaisesti vaikuttavat selittävät muuttujat. Nämä tulee tutkia vähintään tuoteryhmäkohtaisesti.

7.5 Oman oppimisen arviointi

Vastasin kehitys- ja opinnäytetyöprojektista projektipäällikön roolissa ja tutkimuksen suunnittelusta, toteutuksesta ja raportoinnista opinnäytetyön tekijänä. Aloitin elokuussa 2019 tekemään toimeksiantajan eli Haaga-Helia ammattikorkeakoulun kautta kehittämisprojektia kohdeyritykselle. Kehittämis- ja opinnäyteprojekti saatettiin päätökseen toukokuussa 2020.

Kiinnostuin heti big datasta, kun kuulin ensimmäisen kerran sen hyödyntämisestä liiketoiminnassa kansainvälisessä konferenssissa Frankfurtissa vuonna 2013. Kuusi vuotta myöhemmin pääsin mukaan Big Data – Big Business -projektiin, jossa pääsin kehittämään opinnäytetyönä massadatan hyödyntämistä kansainvälisen yrityksen liiketoiminnan näkökulmasta. Pääsin käyttämään projektissa aikaisempaa osaamistani liiketaloudesta, digitalisaatiosta ja markkinoinnista sekä syventämään osaamistani liiketoiminta- ja data-analytiikassa sekä digitaalisessa markkinoinnissa tietojärjestelmäosaamisen -koulutusohjelmassa Haaga-Helia ammattikorkeakoulussa. Perehdyin myös Google Trends-, Ubersuggest- ja IBM SPSS -työkaluihin sekä lukuihin teorioihin muun muassa Kotlerin ja Kellerin ostopäätösprosessiin.

Mielestäni onnistuin mallintamaan hyvin ostopäätösprosessin tiedon etsimisen ja ostopäätösajankohdan vaiheet kohdeyrityksen liiketoiminnan näkökulmasta hyödyntämällä Google Trends -indeksidataa ja mallinnusta varten poimittua myyntidataa. Sen sijaan myynnin ennustemalli, jossa hyödynnettiin historiadataa myynneistä Google Trends -indeksidatan ja mallinnusdatan avulla tuotti haasteita, koska saatavilla ollut data ei ollut riittävän tarkkaa. Lisäksi työ paisui tavanomaista opinnäytetyötä huomattavasti laajemmaksi. Tutkimuksesta olisi kannattanut karsia jo alkuvaiheessa ennustemalli pois, mutta työn mitasuhteet selkeytyivät vasta sitä tehdessä. Toisaalta aihe kiinnosti minua, ja syvensi ammattitaitoani ennustamisessa, joten päätin pitää mallin mukana kokonaisuudessa. Olin vuorovaikutuksessa kohdeyrityksen kanssa koko prosessin ajan, ja saadun palautteen perusteella varmistin, että tutkittavat asiat hyödyttävät yrityksen liiketoimintaa.

Opin tutkimusta tehdessäni paljon kuluttajaymmärrystä lisäävistä datatyypeistä ja -muodoista, datan keräämis- ja analysointimenetelmistä sekä hyödyntämisestä liiketoiminnassa. Perehdyin kohdeyrityksen toimintaympäristöön ja liiketoimintaan, jotta pystyin kehittämään mallit, joilla pääsin tutkimaan ostopäätösprosessin päivittäiselintarvikkeiden hakuajankohtia, Google-hakujen ja ostamisen välistä viivettä sekä myynnin ennustamista. Kehitin myös uudenlaisen, järjestelmällisen tavan yhdistää data-analytiikkaan hakuko-

neanalytiikalla selvitetty kuluttajien käyttämät hakutermiit. Opin soveltamaan erilaisia menetelmiä käytäntöön ja yhdistelemään niitä sekä käyttämään erilaisia työkaluja datan keräämisessä ja analysoinnissa. Ennen kaikkea oivalsin, että jokaisen yrityksen kannattaa laatia datastrategia ja miettiä, miten massadatan avulla voi kehittää uusia liiketoimintamalleja ja kasvattaa liikevaihtoa, tehdä parempia päätöksiä ja lisätä ymmärrystä kuluttajista, markkinoista ja kilpailijoista. Koska datan määrä lisääntyy nopeasti, myös osaamistarpeet ja kehittyminen edellyttävät asioiden jatkuvaa seuranta kilpailukyyn parantamiseksi. Matka kannattaa aloittaa heti tänään.

Lähteet

Aalto, K. 2018. Elintarvikkeiden kulutus kotitalouksissa vuonna 2016 ja muutokset vuosisista 2012, 2006 ja 1998. Kuluttajatutkimuskeskus 2018:80. Valtiotieteellisen tiedekunnan julkaisuja. Helsingin yliopisto. Helsinki. Luettavissa: https://helda.helsinki.fi/bitstream/handle/10138/235324/Elintarvikkeiden_kulutus_kotitalouksissa_2016...pdf?sequence=1&isAllowed=y. Luettu: 5.3.2020.

Ali-Yrkkö, J., Mattila, J., Pajarinen, M. & Seppälä, T. 2019. Digibarometri 2019. Digi tulee, mutta riittävätkö resurssit? Taloustieto Oy. Helsinki. Luettavissa: <http://www.digibarometri.fi>. Luettu: 16.4.2020.

Alko. 2020. Alkon myynti tuoteryhmittäin, suppea raportti. Luettavissa: <https://www.alko.fi/alko-oy/uutishuone/myyntitilastot/myynti/tuoteryhmittain.suppea>. Luettu: 2.5.2020.

Alko. 2019. Alkon myynti maakunnittain. Luettavissa: <https://www.alko.fi/alko-oy/uutishuone/myyntitilastot/myynti/maakunnittain.kaikki-maakunnat>. Luettu: 27.11.2019.

Alkoholilaki 1102/2017. Luettavissa: <https://www.finlex.fi/fi/laki/smur/2017/20171102>. Luettu: 2.5.2020.

Anon. 8.10.2019. Kehitysjohdaja. Kohdeyritys. Haastattelu. Helsinki.

Antikainen, J., Eskelinen, J., Koski, H., Niemi, T., Pajarinen, M., Pyykkönen, S. & de Vries, M. 2016. Massadatasta liiketoimintaa ja tehokkaita julkisia palveluja. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 16/2016. Helsinki. Luettavissa: https://www.etla.fi/wp-content/uploads/vnk_raportti_2016_16-1.pdf. Luettu: 16.4.2020.

Azhari, J. E. & Bennett, D. 1.9.2015. Alumna & Associate Professor. Omni-channel customer experience: An investigation into the use of digital technology in physical stores and its impact on the consumer's decision-making process. London South Bank University. Seminaariesitys. Lontoo. Luettavissa: <https://openresearch.lsbu.ac.uk/item/87613>. Luettu: 27.3.2020.

Backlinko 2020. SEO Tools and Software. Ubersuggest. Luettavissa: <https://backlinko.com/hub/seo/ubersuggest>. Luettu: 26.4.2020.

- Baraka, J. 2019. Iltasanomat. Jouluku on Alkoissa vuoden kiireisintä aikaa – näitä juomia sieltä haetaan juuri nyt eniten. Luettavissa: <https://www.is.fi/ruokala/ajankohtaista/art-2000006339983.html>. Luettu: 24.2.2020.
- Bergström, S. & Leppänen, A. 2015. Yrityksen asiakasmarkkinointi. 16., uudistettu painos. Edita Publishing Oy. Helsinki.
- Bhat, A. 29.2.2020. Ubersuggest Review 2020/Features, Usage & Pricing. Blogi. Luettavissa: <https://www.jinconnect.com/ubersuggest-review/>. Luettu: 7.4.2020.
- Boone, T., Ganeshan, R. & Hicks, R. L. 2015. Incorporating Google Trends Data Into Sales Forecasting. Foresight. Luettavissa: <https://static1.squarespace.com/static/5b9e942a8f5130f854dbef81/t/5be9a6208985835f898e3a0f/1542039073203/incorporating-google-trends-data-into-sales-forecasting.pdf>. Luettu: 10.2.2020.
- Bruun, M. & Mägi, M. 2020. Yritysten dataa tarvitaan tilastojen pohjaksi. Helsingin Sanomat. Mielipide. Luettavissa: <https://www.hs.fi/mielipide/art-2000006370096.html>. Luettu: 13.1.2020.
- Bynum, M. 2019. Ubersuggest: The Ultimate Tool for Higher Google Rankings. binPress. Luettavissa: <https://www.binpress.com/ubersuggest/>. Luettu: 7.4.2020.
- Choi, H. & Varian, H. 2012. Predicting the present with Google Trends. Economic Record, 88, 1, s. 2-9.
- Choi, H. & Varian, H. 2011. Predicting the Present with Google Trends. Google. Luettavissa: <http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf>. Luettu: 10.2.2020.
- Choi, H. & Varian, H. 2009. Predicting the Present with Google Trends. Google. Luettavissa: https://static.googleusercontent.com/media/www.google.com/fi/googleb-logs/pdfs/google_predicting_the_present.pdf. Luettu: 10.2.2020.
- Clear Channel 2020a. Suunnitteluohjeita. Luettavissa: <https://clearchannel.fi/kampanja-suunnittelu/>. Luettu: 23.4.2020.

Clear Channel 2020b. Tavoita asiakkaasi ostopäätöksen hetkellä. Helsinki. Luettavissa: <https://info.clearchannel.fi/clear-channel-store-digital-myymalamedia-opas?hsCtaTracking=0c4818ff-ca2c-4c85-b4af-f89f3cced031%7C7a077e17-f75b-408c-a3d7-3deb3abf13b4>. Luettu: 28.3.2020.

Crestodina, A. 17.7.2017. Competitor Analysis Tools: 15 Quick Ways to Compare Websites. Blogi. Luettavissa: <https://www.orbitmedia.com/blog/website-competitive-analysis-tools/>. Luettu: 16.9.2019.

Davenport, T. H. & Harris, J. G. 2017. Competing on analytics: the new science of winning. Harvard Business Review Press. Boston.

Elonen, P. 2020. Ruuan verkkokauppa räjähti käsiin – kaksi keinoa löytää vapaita kuljetusaikoja. Helsingin Sanomat. Luettavissa: <https://www.hs.fi/talous/art2000006458195.html>. Luettu: 20.3.2020.

Elzinga, D. 2017. Driving business growth by zeroing in on the consumer decision journey. Podcast transcript. McKinsey & Company. Luettavissa: <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/driving-business-growth-by-zeroing-in-on-the-consumer-decision-journey>. Luettu: 26.3.2020.

Euroopan parlamentin ja neuvoston asetus (EU) 2016/679 luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta. Annettu 27.4.2016.

FIAM 2020a. Finnish Internet Audience Measurement. comScoren yleisömittaustuotteet lyhyesti. Luettavissa: <https://fiam.fi/2018/02/01/comscoren-metodologia-ja-yleisomittaus-tuotteet-pahkinankuoressa/>. Luettu: 4.5.2020.

FIAM 2020b. Finnish Internet Audience Measurement. FIAM tuottaa Suomen viralliset digitaalisten yleisöjen luvut sivustoista ja sovelluksista. Luettavissa: <https://fiam.fi/>. Luettu: 31.3.2020.

Finanssiala ry 2019. Massadata ja data-analytiikka. Luettavissa: <http://www.finanssialalle.fi/opintomateriaalit/finanssialan-perusteet/innovaatiot/massadata-ja-data-analytiikka.html>. Luettu: 25.11.2019.

Finder 2020. Kaikki olennainen suomalaisista yrityksistä. Yhteystiedot, päättäjät, taloustiedot sekä palvelut. Luettavissa: <https://www.finder.fi/>. Luettu: 5.4.2020.

Follet, M. 2020. Huomio: median yhteinen valuutta. Lumen Research & JCDecaux. Luettavissa: <https://www.jcdecaux.fi/Article/huomio-median-yhteinen-valuutta>. Luettu: 6.4.2020.

Fortune Business Insights 2020. Big Data Technology Market Size, Share & Industry Analysis, By Offering (Solution, Services), By Deployment (On-Premise, Cloud, Hybrid), By Application (Customer Analytics, Operational Analytics, Fraud Detection and Compliance, Enterprise Data Warehouse Optimization, Others), By End Use Industry (BFSI, Retail, Manufacturing, IT and Telecom, Government, Healthcare, Utility, Others) and Regional Forecast, 2019–2026. Luettavissa: <https://www.fortunebusinessinsights.com/industry-reports/big-data-technology-market-100144>. Luettu: 16.4.2020.

Google 2020a. Google Search Statistics. Luettavissa: <https://www.internetlivestats.com/google-search-statistics/>. Luettu: 3.5.2020.

Google. 2020b. Google Trends. Trendit Ohjeet. Luettavissa: <https://support.google.com/trends/answer/4359550?hl=fi>. Luettu: 10.2.2020.

Google Trends. 2020. Kiinnostus alialueittain > Info.

Haaga-Helia ammattikorkeakoulu 2020. Tutkimuksen luotettavuus. Moodle. Intranet. Luettu: 10.3.2020.

Haaga-Helia ammattikorkeakoulu 2019a. Big Data – Big Business. Hankkeet. Luettavissa: <http://www.haaga-helia.fi/fi/big%20data>. Luettu: 27.9.2019.

Haaga-Helia ammattikorkeakoulu 2019b. Organisaatio. Luettavissa: <http://www.haaga-helia.fi/fi/haaga-heliasta/organisaatio?userLang=fi>. Luettu: 27.9.2019.

Haaga-Helia ammattikorkeakoulu 2019c. Opinnäytetyön eettiset ohjeet. MyNet. Intranet. Luettu: 30.9.2019.

Haraldsen, G., Amdam, S. & Zhang, L.-C. 2014. Towards an integrated Consumer Expenditure Survey – Combining Multi-mode Data Collection and Big Data Extracts. Luettavissa: https://ec.europa.eu/eurostat/cros/system/files/Haraldsen-etal_NTTS2015abstract-20140715-unblinded_0.pdf. Luettu: 26.3.2020.

Hatakka, T. 2015. Soveltuvatko kaupan kanta-asiakasrekisterit kulutuksen tilastointiin? Tilastokeskus. Luettavissa: <https://www.stat.fi/tietotrendit/artikkelit/2015/soveltuvatko-kaupan-kanta-asiakasrekisterit-kulutuksen-tilastointiin/>. Luettu: 26.3.2020.

Helsingin yliopiston Kalenteripalvelut Oy 2018. Kalenteri vuodelle 2019. Yliopiston almanakkatoimisto. Luettavissa: <https://almanakka.helsinki.fi/fi/julkaisut/kalenteri-vuodelle-2019.html>. Luettu: 13.2.2020.

Helsingin yliopiston Kalenteripalvelut Oy 2017. Kalenteri vuodelle 2018. Yliopiston almanakkatoimisto. Luettavissa: <https://almanakka.helsinki.fi/fi/27-suomi/julkaisut/476-kalenteri-vuodelle-2018.html>. Luettu: 13.2.2020.

Helsingin yliopiston Kalenteripalvelut Oy 2014. Kalenteri vuodelle 2015. Yliopiston almanakkatoimisto. Luettavissa: <https://almanakka.helsinki.fi/fi/julkaisut/ajankohtaista/27-suomi/julkaisut/377-kalenteri-vuodelle-2015.html>. Luettu: 13.2.2020.

Herrmann, N. 19.10.2018. The Digital Analytics Guide. quintly Blog. Luettavissa: <https://www.quintly.com/blog/using-digital-analytics>. Luettu: 21.10.2020.

Holmes, E. E., Scheuerell, M. D. & Ward, E. J. 2020. Applied Time Series Analysis for Fisheries and Environmental Sciences. Correlation within and among time series. Cross-correlation function (CCF). Northwest Florida State College. Luettavissa: <https://nwfsc-timeseries.github.io/atsa-labs/sec-tslab-correlation-within-and-among-time-series.html>. Luettu: 26.4.2020.

HubSpot 10.5.2017. How to Optimize Your YouTube Videos. YouTube-video. Luettavissa: https://www.youtube.com/watch?time_continue=13&v=h-HIBPx8ek4. Luettu: 26.4.2020.

IBM 2020a. Time Series: Cross-Correlations. How-to Guide for IBM® SPSS® Statistics Software. Luettavissa: <https://methods.sagepub.com/dataset/howtoguide/time-series-cross-correlations-in-us-feedgrains-1876-2015>. Luettu: 13.3.2020.

IBM 2020b. Ohjelmistot – Statistics. Ohjelmistot & ratkaisut. Luettavissa: <https://spss.fi/ohjelmistot-ratkaisut>. Luettu: 19.4.2020.

IBM 2012. CRISP-DM Help Overview. Luettavissa: https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm. Luettu: 5.4.2020.

IltaSanomat 2015. Päivän lämpimin hetki on nyt - tässä syy. Luettavissa: <https://www.is.fi/kotimaa/art-2000000955765.html>. Luettu: 27.4.2020.

JCDecaux 2020a. Ulkomainonnan rooli kuluttajan ostopolulla. Luettavissa: <https://www.jcdecaux.fi/Article/ulkomainonnan-rooli-kuluttajan-ostopolulla>. Luettu: 28.3.2020.

JCDecaux 2020b. Kuluttajan ostopolku ja tavoittaminen päätöksen hetkellä. Luettavissa: <https://www.jcdecaux.fi/uutiset/kuluttajan-ostopolku-ja-tavoittaminen-paatoksen-hetkella>. Luettu: 28.3.2020.

JCDecaux 2020c. Data auttaa ulkomainontaa tavoittamaan parhaan yleisön. Luettavissa: <https://www.jcdecaux.fi/uutiset/data-auttaa-ulkomainosta-tavoittamaan-parhaan-yleison>. Luettu: 28.3.2020.

Jukarainen, M. 20.11.2019. Avainsanatutkimus – Näin selvität parhaat avainsanat helposti! Blogi. Luettavissa: <https://mikaeljukarainen.com/avainsanatutkimus/>. Luettu: 19.4.2020.

Järvinen, J. & Brännare, S. 2020. Lidl ottaa käyttöön oman versionsa bonuskortista – ja kaantuvatko asiakkaat kahteen kastiin? Yle. Luettavissa: <https://yle.fi/uutiset/3-11217248>. Luettu: 6.3.2020.

Kajantie, M. 18.11.2019. Yliaktuaari. Tilastokeskus. Henkilökohtainen tiedonanto. Helsinki.

Kalenteripalvelu Oy 2016. Akava-kalenteri 2017.

Kalenteripalvelu Oy 2015. Akava-kalenteri 2016.

Kananen, J. & Makkonen, T. (toim.) 2015. Opinnäytetyön kirjoittajan opas. Näin kirjoitan opinnäytetyön tai gradun alusta loppuun. Jyväskylän ammattikorkeakoulun julkaisuja - sarja nro 202. Jyväskylän ammattikorkeakoulu. Jyväskylä.

Karadimitriou, S. M. & Marshall, E. 2020. Outliers, Durbin-Watson and interactions for regression in SPSS. Statstutor Community Project. University of Sheffield. Sheffield. Luettavissa: https://www.sheffield.ac.uk/polopoly_fs/1.531431!/file/MASHRegression_Further_SPSS.pdf. Luettu: 30.4.2020.

Kaushik, A. 2007. Web Analytics: an hour a day. Sybex cop. Indianapolis.

Kesko Oyj 2020a. K-Ruoka- ja verkkokaupparekisteri. Luettavissa: https://tietosuoja.kesko.fi/category/k_ruoka_ja_verkkokauppa. Luettu: 6.3.2020.

Kesko Oyj 2020b. Tietosuojaseloste, K-plussan asiakasrekisteri. Luettavissa: <https://www.plussa.com/Mika-on-K-Plussa/Sopimusehdot/Rekisteriselosteen/>. Luettu: 6.3.2020.

Kinski, A. 2016. Google Trends as Complementary Tool for New Car Sales Forecasting: A Cross-Country Comparison along the Customer Journey. University of Twente. Berlin. Luettavissa: <https://essay.utwente.nl/70462/>. Luettu: 28.11.2019.

Kivelä, L. 24.6.2019. Tunnetko ostoprosessin eri vaiheet? Kasvua sisältömarkkinoinnilla. Grapevinen blogi. Luettavissa: <https://grapevine.fi/tunnetko-ostoprosessin-eri-vaiheet-kasvua-sisaltomarkkinoinnilla/>. Luettu: 19.4.2020.

Klemola, A. 2019. Essays on irrational investors' behavioral biases and pricing efficiency. Vaasan yliopisto. Vaasa. Luettavissa: <https://osuva.uwasa.fi/handle/10024/9901>. Luettu: 26.3.2020.

Kotler, P. & Keller, K. L. 2016. Marketing Management. 15. global edition. Pearson Education. Courier. Kendallville.

Kukkonen, L. 2018. Ulkomainonnan ohjelmallinen ostaminen tuli Suomeen – K-ryhmä houkuttelee opiskelijoita kanta-asiakkaiksi. Markkinointi & Mainonta. Luettavissa: <https://www.marmai.fi/uutiset/ulkomainonnan-ohjelmallinen-ostaminen-tuli-suomeen-k-ryhma-houkuttelee-opiskelijoita-kanta-asiakkaiksi/ab560c71-bbdb-35d0-b87d-87f5b3e9ddf8>. Luettu: 6.4.2020.

Kuulas Helsinki/IPG Mediabrands Finland 2019. Retail Buying Study. Tracking the retail scene since 2011. Luettavissa: <https://www.retailbuyingstudy.fi/>. Luettu: 27.3.2020.

Lampela, R. 2018. Tässä ovat maailman arvokkaimmat yritykset – kiinalaiset ylsivät kymmenen kärkeen ensimmäisen kerran. Tivi. Luettavissa: <https://www.tivi.fi/uutiset/tassa-ovat-maailman-arvokkaimmat-yritykset-kiinalaiset-ylsivat-kymmenen-karkeen-ensimmais-kerran/aa7d4271-19a7-3e61-b659-5f6b8e2e8788>. Luettu: 27.4.2020.

Lindberg, S. 22.1.2020. Senior Specialist Programmatic Marketing -asiantuntija. Kesko K Digital. Henkilökohtainen tiedonanto. Helsinki.

Linnake, T. 2020. Suosittu virustutka myy käyttäjätietoja – jopa katsotut pornovideot paljastuvat. Iltasanomat. Digitoday. Luettavissa: <https://www.is.fi/digitoday/tietoturva/art-2000006388076.html>. Luettu: 31.3.2020.

Marr, B. 2016. Key Business Analytics. Pearson Education. Harlow.

Mauranen, K. 2020. Epälineaarisista regressioanalyyseistä. Opetusmateriaali verkossa. UEF: Itä-Suomen yliopisto. Luettavissa: <http://cs.uef.fi/~mauranen/spss/jatko/regressio-analyysi10.html>. Luettu: 22.3.2020.

Mediateko 2020. Ulkomainonnan perusteet. Luettavissa: <https://www.mediateko.fi/ulkomainonta>. Luettu: 28.3.2020.

Mellin, I. 2006. Tilastolliset menetelmät: Lineaarinen regressioanalyysi. Aalto-yliopisto. Luettavissa: <https://math.aalto.fi/opetus/sovtoda/oppikirja/Regranal.pdf>. Luettu: 30.4.2020.

Mueller, J. P. & Massaron, L. 2018. Artificial intelligence for dummies. John Wiley & Sons. Hoboken.

Nielsen 2020. Liity osaksi Nielsenin kuluttajaneelaa, meitä on jo lähes 5000: erikokoisia ja -ikäisiä talouksia! Luettavissa: https://www.nielsenonline.com/?cpid=1A0FC5DE-&transaction_id=102affb57b1c98342f84d656755f25. Luettu: 9.1.2020.

Nielsen 2019. Päivittäistavarakaupan myymälärekisteri 2018. Luettavissa: <https://www.nielsen.com/fi/fi/press-releases/2019/grocery-store-register-2018/>. Luettu: 6.12.2019.

Norton, D. W. & Pine II, B. J. 2013. Using the customer journey to road test and refine the business model. *Strategy & Leadership*, 41, 2, s. 12–17.

Nummenmaa, L. 2011. Käyttäytymistieteiden tilastolliset menetelmät. 3., uudistettu laitos. Kustannusosakeyhtiö Tammi. Helsinki.

Oberlo 2020. Search Engine Market Share in 2019. Luettavissa: <https://www.oberlo.com/statistics/search-engine-market-share>. Luettu: 6.3.2020.

Ojasalo, K., Moilanen, T. & Ritalahti, J. 2014. Kehittämistyön menetelmät: Uudenlaista osaamista liiketoimintaan. 3. uudistettu painos. Sanoma Pro. Helsinki.

Outdoor Impact 2020. Mitä VAC tarkoittaa? Luettavissa: http://www.outdoorimpact.fi/index.php?loc_id=32. Luettu: 23.4.2020.

Paljakka, E. 2019. Suomen ostetuin eines on jo 62-vuotias - jopa Henri Alén antaa kiitosta. *Ilta-lehti*. Luettavissa: <https://www.iltalehti.fi/ruoka-artikkelit/a/4b35098b-a3a3-4a55-ad62-cea52e028177>. Luettu: 6.4.2020.

Palokangas, P. 2019. Suomalainen aloittaa ostopolkunsa yhä useammin netistä: ”Kun vertailemme tuloksia Amazon-maihin, muutos on vasta alussa.” *Markkinointi & Mainonta*. Luettavissa: <https://www.marmai.fi/uutiset/suomalainen-aloittaa-ostopolkunsa-yha-useammin-netista-kun-vertailemme-tuloksia-amazon-maihin-muutos-on-vasta-alussa/f488b2d8-e144-4460-acda-3cda8e34b112>. Luettu: 26.4.2020.

Piippo, M. 2016. Ilmaiset SEO-työkalut TOP 25 – näillä säästät tuhansia euroja! *Hopkins*. Luettavissa: <https://www.hopkins.fi/artikkelit/ilmaiset-seo-tyokalut-top-25/>. Luettu: 26.4.2020.

Prantl, D. & Prantl, M. 2018. Website traffic measurement and rankings: competitive intelligence tools examination. *International Journal of Web Information Systems*, 14, 4, s. 423–437.

Päivittäistavarakauppa ry 2020. Suomen päivittäistavaramarkkinat. Luettavissa: <https://www.ptv.fi/kaupan-toiminta/paivittaeistavaramarkkinat-suomessa/>. Luettu: 15.4.2020.

Raijas, A. 25.2.2016. Mitä kuluttaja on? Kulma. Kilpailu- ja kuluttajaviraston kuluttajakasvatuksen ja -osaamisen blogi. Luettavissa: <https://kulma.kkv.fi/2016/02/25/mita-kuluttaja-on/>. Luettu: 10.5.2020.

Raivio, L. 2018. Hyötyä omaan koriin. Yhteishyvä 2, s. 88–91.

Relevant 29.11.2019. Mitä ovat 1st, 2nd ja 3rd party data ja miten niitä voidaan hyödyntää markkinoinnissa? Blogi. Luettavissa: <https://relevant.fi/blogi/mita-ovat-1st-2nd-ja-3rd-party-data-ja-miten-niita-voidaan-hyodyntaa-markkinoinnissa/>. Luettu: 7.3.2020.

Riipinen, M. 2019. Ulkomainonta. Massamedian tavoittavuus, verkkomedian ketteryys. Dagmar. Luettavissa: <https://www.dagmar.fi/digitaalinen-markkinointi/ulkomainonta-massamedian-tavoittavuus-verkkomedian-ketteryys/>. Luettu: 28.3.2020.

Rissanen, A. 2020. Lidlin kansainvälinen etuohjelma Lidl Plus testissä Suomessa – lanseerataan kevään aikana. Luettavissa: <https://news.cision.com/fi/lidl-suomi/r/lidlin-kansainvalinen-etuohjelma-lidl-plus-testissa-suomessa---lanseerataan-kevaan-ai-> kana,c3049370. Luettu: 6.3.2020.

Rouhiainen, L. 2019. Digital marketing in Modern Business. Verkkokurssi. Haaga-Helia ammattikorkeakoulu.

Rouse, M. 2017. Data Analytics (DA). Tech Target. Luettavissa: <https://searchdata-management.techtarget.com/definition/data-analytics>. Luettu: 26.3.2020.

Saaranen, P. 10.3.2020. Yliopettaja. Haaga-Helia ammattikorkeakoulu. Henkilökohtainen tiedonanto. Helsinki.

Saaranen, P. 18.3.2020. Yliopettaja. Haaga-Helia ammattikorkeakoulu. Henkilökohtainen tiedonanto. Helsinki.

Saaranen, P. 2018. Tilastolliset menetelmät. Moniste.

Saaranen, P. 2017. Kyselytutkimuksen analysointi SPSS-ohjelmalla. Moniste.

Salminen, J. 2.4.2015. Opettaja. Digitaalinen analytiikka. Turun kauppakorkeakoulu. Luento. Turku. Luettavissa: <https://www.slideshare.net/jonis12/digitaalinen>. Luettu: 26.3.2020.

Salo, I. 2014. Big Data & pilvipalvelut. Docendo Oy. Jyväskylä.

Sanastokeskus TSK 2020. Erikoisalojen sanastojen ja sanakirjojen kokoelma. TEPA-termipankki. Luettavissa: <http://www.tsk.fi/tepa/fi/haku/spatiotemporaalinen%20analyysi>. Luettu: 31.3.2020.

Sharda, R., Delen, D., Turban, E., Aronson, J. E., Ting-Peng, L. & King, D. 2018. Business intelligence, analytics, and data science: a managerial approach. 4. global edition. Pearson. Harlow.

Sipinen, H. 2018. IL selvitti: Näin alkoholin tilaaminen ulkomailta onnistuu - nuhteeton ostaja ei juuri hyödy rahallisesti. Iltalehti. Luettavissa: <https://www.iltalehti.fi/kotimaa/a/201807292201093207>. Luettu: 2.5.2020.

SOK 2020. Suomen Osuuskauppojen keskuskunta. S-ryhmän asiakasomistaja- ja asiakasrekisteri. Tietosuojaseloste. Luettavissa: <https://www.s-kanava.fi/asiakasomistaja/artikkeli/s-ryhman-asiakasomistaja-ja-asiakasrekisteri/580dr26haWkdjYWimKQkVM>. Luettu: 6.3.2020.

Sponder, M. & Khan, G. F. 2018. Digital Analytics for Marketing. Routledge. New York.

Statcounter 2020a. Search Engine Market Share WorldWide – March 2019–March 2020. Luettavissa: <https://gs.statcounter.com/search-engine-market-share>. Luettu: 26.4.2020.

Statcounter 2020b. Search Engine Market Share in Finland – March 2019–March 2020. Luettavissa: <https://gs.statcounter.com/search-engine-market-share/all/finland>. Luettu: 26.4.2020.

Stephens-Davidowitz, S. & Varian, H. 2015. A Hands-on Guide to Google Data. Google. Luettavissa: <http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf>. Luettu: 10.2.2020.

Sticky Media 2020. The Importance of Google Analytics – Get the Facts About Your Website. Luettavissa: <https://www.stickymedia.com/blog/the-importance-of-digital-analytics/>. Luettu: 25.3.2020.

Sulopuisto, O. 2018. Yhden asian medioita. Suomen Lehdistö. Luettavissa: <https://suomenlehdisto.fi/yhden-asian-medioita/>. Luettu: 31.3.2020.

Suojanen, J. 13.7.2016. Suomen Digimarkkinointi Oy. Mitä hakukoneoptimointi (SEO) on? Blogi. Luettavissa: <https://www.digimarkkinointi.fi/blogi/mita-hakukoneoptimointi-seo-on>. Luettu: 6.12.2019.

Susanto, H., Sucahyo, Y. G., Ruldeviyani, Y. & Gandhi, A. 17.1.2019. Analysis of Factors that Influence Purchase Intention on Omni-channel Services. Universitas Indonesia. Seminarisitys. Jakarta. Luettavissa: <https://ieeexplore.ieee.org/document/8618249>. Luettu: 27.3.2020.

Taanila, A. 22.12.2019. Aikasarjaennustaminen 1. Akin menetelmäblogi. Luettavissa: <https://tilastoapu.wordpress.com/2018/08/30/aikasarjaennustaminen-1/>. Luettu: 20.3.2020.

Taanila, A. 21.4.2019. Tunnuslukuja. Akin menetelmäblogi. Luettavissa: <https://tilastoapu.wordpress.com/2011/10/19/tunnuslukuja/>. Luettu: 20.3.2020.

Taanila, A. 19.4.2019a. Virhemarginaali ja luottamusväli. Akin menetelmäblogi. Luettavissa: <https://tilastoapu.wordpress.com/virhemarginaali-2/>. Luettu: 20.3.2020.

Taanila, A. 19.4.2019b. Keskiarvon virhemarginaali. Akin menetelmäblogi. Luettavissa: <https://tilastoapu.wordpress.com/2011/10/27/9-keskiarvon-virhemarginaali/>. Luettu: 20.3.2020.

Taanila, A. 17.4.2019. Korrelaatio ja sen merkitsevyys. Akin menetelmäblogi. Luettavissa: <https://tilastoapu.wordpress.com/2011/11/01/10-korrelaatio-ja-sen-merkitsevyys/>. Luettu: 20.3.2020.

Taanila, A. 31.3.2019. Mittaamisen luotettavuus. Akin menetelmäblogi. Luettavissa: <https://tilastoapu.wordpress.com/2012/03/14/mittaamisen-luotettavuus/>. Luettu: 20.3.2020.

Taanila, A. 2019a. Data-analytiikka. Kurssimateriaali. Haaga-Helia ammattikorkeakoulu. Helsinki. Luettu: 20.3.2020.

Taanila, A. 2019b. Datan esittäminen ja kuvailu. Taanila. Helsinki. Luettavissa: <http://taanila.fi/kuvailu.pdf>. Luettu: 20.3.2020.

Taanila, A. 2015. Toimitusketjun hallinnan työkaluja. Taanila. Helsinki. Luettavissa: <http://myy.haaga-helia.fi/~taaak/s/toimitusketju.pdf>. Luettu: 19.3.2020.

Taanila, A. 2010. Lineaariset regressiomallit. Taanila. Helsinki. Luettavissa: <http://myy.haaga-helia.fi/~taaak/m/regressio.pdf>. Luettu: 19.3.2020.

Terveysten- ja hyvinvoinnin laitos 2019. Alkoholijuomien matkustajatuontikysely (syyskuu 2018–elokuu 2019). Tilastoraportti 35/2019. Helsinki. Luettavissa: <https://www.julkari.fi/handle/10024/138663>. Luettu: 2.5.2020.

Tietosuojavaltuutetun toimisto 2020. Usein kysyttyä EU:n tietosuojasetuksesta. Luettavissa: <https://tietosuoja.fi/gdpr>. Luettu: 6.3.2020.

Tilastokeskus 2020a. Tee oma karttaesitys. Luettavissa: <https://www.tilastokeskus.fi/tup/aluetietoa/tee-oma-karttaesitys.html>. Luettu: 5.3.2020.

Tilastokeskus 2020b. Avoin data. Luettavissa: <https://www.stat.fi/org/avoindata/index.html>. Luettu: 5.4.2020.

Tilastokeskus 2019a. Yksilöllisen kulutuksen käyttötarkoituksen mukainen luokitus COICOP. Luettavissa: <https://www.stat.fi/meta/luokitukset/coicop/011-2013/0114.html>. Luettu: 28.11.2019.

Tilastokeskus 2019b. Pääluokat – Yksilöllisen kulutuksen käyttötarkoituksen mukainen luokitus COICOP. Maitotuotteet, juusto ja kananmunat. Luettavissa: <https://www.stat.fi/meta/luokitukset/coicop/011-2013/0114.html>. Luettu: 28.11.2019.

Tilastokeskus 2019c. Pääluokat – Yksilöllisen kulutuksen käyttötarkoituksen mukainen luokitus COICOP. Valmisateriat. Luettavissa: <https://www.stat.fi/meta/luokitukset/coicop/011-2013/0114.html>. Luettu: 28.11.2019.

Tilastokeskus 2019d. Pääsiäistilastoja 2019. Luettavissa: https://www.stat.fi/tup/tilastokirjasto/paasiaistilastoja_2019.html. Luettu: 28.11.2019.

Tilastokeskus 2019e. Vapputilastoja 2019. Luettavissa: https://www.stat.fi/tup/tilastokirjasto/vapputilastoja_2019.html. Luettu: 28.11.2019.

Tilastokeskus 2019f. Suomen kesä tilastojen valossa 2019. Luettavissa: http://www.stat.fi/tup/tilastokirjasto/kesatilastot_2019. Luettu: 28.11.2019.

Tilastokeskus 2019g. Maakuntaluokitus. Luettavissa: <https://www.tilastokeskus.fi/meta/luokitukset/maakunta/001-2019/index.html>. Luettu: 18.1.2020.

Tilastokeskus 2018a. Joulu tilastojen valossa 2018. Luettavissa: <http://www.stat.fi/tup/tilastokirjasto/joulu-tilastojen-valossa-2018>. Luettu: 28.11.2019.

Tilastokeskus 2018b. Kulutustutkimus 2016. Käyttäjän käsikirja. Helsinki. Luettavissa: http://www.stat.fi/tup/julkaisut/tiedostot/julkaisuluettelo/yksk1_201600_2018_19742-_net.pdf. Luettu: 28.11.2019.

Tuominen, P. 2019. Lidlistä kerrotaan, milloin on pahin ryysis: Jos haluat välttää jouluruuhkan joulukaupassa, älä mene ostoksille tänä ajankohtana. MTV Uutiset. Luettavissa: <https://www.mtvuutiset.fi/makuja/artikkeli/lidlista-kerrotaan-milloin-on-pahin-ryysis-jos-haluat-valttaa-jouluruuhkan-ruokakaupassa-ala-mene-ostoksille-tana-ajankohtana/766-4228#gs.5n1a0l>. Luettu: 6.3.2020.

Turunen, J. 2020. Mitä tapahtuu verkon yksityisyydelle? Helsingin Sanomat. Luettavissa: <https://www.hs.fi/talous/art-2000006384928.html>. Luettu: 26.1.2020.

Ubersuggest 2019. Ubersuggest. Luettavissa: <https://neilpatel.com/ubersuggest/>. Luettu: 10.2.2020.

Valio 2019. Joulukinkku – ohjeet, ideat ja vinkit. Luettavissa: <https://www.valio.fi/joulu/joulukinkku/>. Luettu: 6.12.2019.

Valitut Matkat 14.12.2019. Alkon aukioloajat jouluna 2019. Blogi. Luettavissa: <http://valitutmatkat.blogspot.com/2019/12/alkon-aukioloajat-jouluna-2019.html>. Luettu: 24.2.2020.

Vegaanituotteet.net 2019. Vegaanituotteet. Luettavissa: <http://www.vegaanituotteet.net/>. Luettu: 28.11.2019.

Verhoef, P. C., Kannan, P. K. & Inman, J. J. 2015. From Multi-Channel Retailing to Omni-Channel Retailing. Introduction to the Special Issue on Multi-Channel Retailing. Journal of Retailing 91, 2, s. 174–181.

Wikipedia 2020a. Fast-moving consumer goods (FMCG). Luettavissa: https://en.wikipedia.org/wiki/Fast-moving_consumer_goods. Luettu: 15.4.2020.

Wikipedia 2020b. Asiakas. Luettavissa: <https://fi.wikipedia.org/wiki/Asiakas>. Luettu: 10.5.2020.

Yhteishyvä. 2020. Kinkkusesonki alkaa Lucian päivästä – ”Luomukinkku kannattaa ostaa ajoissa.” Luettavissa. <https://yhteishyva.fi/artikkeli/kinkkusesonki-alkaa-lucian-paivasta-luomukinkku-ka/408ad8MQY0O4w6Uey2YycE>. Luettu: 6.3.2020.

Ylä-Anttila, A. 2019. Tutkimus: Ulkomainonta tavoittaa 97 prosenttia väestöstä – ”Näköjärjestelmämme on erityisen herkkä liikkeelle.” Verkkosivut. 8.2.2019. Markkinointi & mainonta. Luettavissa: <https://www.marmai.fi/uutiset/tutkimus-ulkomainonta-tavoittaa-97-prosenttia-vaestosta-nakojarjestelmamme-on-erityisen-herkka-liikkeelle/259ead92-e7ad-3ad0-a904-7c4cfae1ede0>. Luettu: 28.3.2020.

Liitteet

Liite 1. Keskeiset käsitteet, lyhenteet ja englanninkieliset termit

Aikasarja Tarkasteltavan muuttujan datapisteiden jakso mitattuna ja esitettynä peräkkäisten pisteiden ajanjaksoina, jotka on jaettu tasaisilla aikaväleillä (Sharda 2018)

Aikasarjaennustaminen Ennustemalli, joka luottaa yksinomaan muuttujan aikaisempiin arvoihin arvioidessaan tulevia arvoja (Sharda ym. 2018)

Analytiikka Datan laajaa käyttöä, tilastollista ja kvantitatiivista analysointia, selittäviä ja ennustavia malleja sekä tietoon perustuvaa päätöksentekoa ja toimintaa (Davenport & Harris 2017)

Asiakas Henkilö, joka ostaa hyödykkeen suorittaen maksun sopimuksen mukaan. Maksu voi olla rahaa, tuotteita tai palveluita. Yrityksen ulkopuolinen asiakassuhde voi olla joko suora tai epäsuora (Wikipedia 2020)

Autokorrelaatio Toisistaan riippuvien jäännösten pistejoukossa on satunnaisuudesta poikkeavaa säännönmukaisuutta (Taanila 2010)

Avainsana Hakusana, hakutermi

Beta-kerroin Regressiosuoran kulmakerroin normitetuille arvoille (Saaranen 2018)

Big data Massadata viittaa datan määrään (volume), vaihteluun (variety), vauhtiin (velocity), todenmukaisuuteen (veracity) ja arvoon (value) (Marr 2016)

Brändi Mielikuva tuotteesta

Census-data Selainpohjaisella trafiikkimittauksella kerätty data, josta näkee liikenteen määrän (Kaushik 2007)

COICOP-HBS Eurostatin kansainvälinen yksilöllisen kulutuksen käyttötarkoituksen mukainen luokitus (Tilastokeskus 2019)

CRISP-DM Cross-Industry Standard Process for Data-Mining (IBM 2012), tiedonlouhinnan poikkialainen standardiprosessi muodostuu kuudesta vaiheesta, joka alkaa liiketoiminnan

ja projektin tarpeen hyvästä ymmärtämisestä ja päättyy ratkaisun käyttöönottoon (Sharda ym. 2018)

Data Esimerkiksi nimiä ja numeroita, jotka ovat sellaisenaan merkityksettömiä faktoja (Sharda ym. 2018)

Data-analytiikka Tietojoukkojen sisältämän informaation tarkastelua johtopäätösten tekemiseksi yleensä tietojärjestelmien ja ohjelmien avulla (Rouse 2017)

Digitaalinen analytiikka Kvalitatiivisen ja kvantitatiivisen liiketoiminta- ja kilpailijadatan analysointia, jolla ohjataan asiakkaiden verkkokokemuksen jatkuvaa parantamista kohti toivottua lopputulosta (Sticky Media 2020)

Digitaalinen ulkomainonta DOOH, Digital-Out-Of-Home. Digitaaliset mainospinnat (Mediateko 2020)

Elintarvikkeiden kulutus Tarkoitetaan elintarvikkeita, jotka on ostettu kotikäyttöön kauppoista, kioskeista ja toreilta, ja lisäksi itse omaan käyttöön kasvatettuja puutarha- ja maataloustuotteita, itse kerättyjä marjoja ja sieniä sekä itse pyydettyä riistaa ja kaloja (Aalto 2018)

Ennustaminen Forecasting. Historiatietojen käyttäminen ennakoimaan muuttujan tulevia arvoja (Sharda ym. 2018)

Ennustava analytiikka Predictive analytics. Liiketoiminnan analyttinen lähestymistapa ennustaa esimerkiksi kysyntää, ongelmia tai mahdollisuuksia (Sharda ym. 2018)

Ennustetut arvot Predicted value

Ennustus Prediction. Toimi, jolla kerrotaan tulevasta (Sharda ym. 2018)

Ensimmäisen osapuolen data Kerätään suoraan omalta yleisöltä tai asiakkailta (Sponder & Khan 2018)

FIAM The Finnish Internet Audience Measurement. Online-yleisöjen virallinen mittaus (FIAM 2020b.)

FMCG-tuotteet Fast Moving Consumer Goods. Päivittäistavarat eli nopeasti liikkuvat, kuluttajille suunnatut kulutustavarat, jotka ostetaan usein ruokaostojen yhteydessä (Wikipedia 2020)

F-suhde Mallin sopivuuden tunnusluku (Nummenmaa 2011)

GDPR General Data Protection Regulation (EU:n yleinen tietosuoja-asetus 679/2016/EU)

Google Trends Avoin hakukoneanalytiikkatyökalu, jolla saa tietoa Google-hakukoneella tehdyistä hauista hakutermien ja -kategorian, ajan ja maantieteellisen sijainnin perusteella (Sponder & Khan 2018)

Hajontakuvi Hajontakaavio, Scatter. Kahden jatkuvan muuttujan yhteisjakauman esittämiseen soveltuva kuvi (Saaranen 2018)

Hajontaluvut Kuvataan muuttujan arvojen jakautumista esimerkiksi keskiarvon ympärillä, jolloin niistä käytetään hajonta-termiä (Nummenmaa 2011)

Hakukone Ohjelma, joka löytää ja listaa verkkosivut, jotka täsmäävät käyttäjän valitsemiin kriteereihin (Sharda ym. 2018)

Hakukoneanalytiikka Hakukoneoptimointi (SEO, Search Engine Optimazion) ja hakutrendien analysointi (Sponder & Khan 2018)

Hakukonedata Hakukoneiden hakutulokset (Kaushik 2007)

Hakutrendien analysointi Ymmärrys, millä avainsanoilla ihmiset tekevät hakuja tiettyä ajanjaksona (Sponder & Khan 2018)

Huipukkuusluku Kurtosis. Kuvaa poikkeamaa normaalisti jakautuvasta kuvaajasta: positiivinen arvo huipukkuutta ja negatiivinen arvo laakeaa jakaumaa (Saaranen 2018)

IBM SPSS Statistics Statistical Package for Social Sciences. Tilasto-ohjelma numeroaineistojen analysointiin (IBM 2020)

ISP Internet Service Providers. Mittaustapa, joka käyttää anonymia erilaisilta verkkopalveluiden tarjoajilta kaapattua dataa (Kaushik 2007)

Jäännöstermi Jäännös, jäännösvirhe, virhetermi, virhearvio, residuaali, residual. Kuvaa regressiomallissa esiintyvää virhettä (Saaranen 2017)

Kertoimet Coefficients

Keskeinen raja-arvolause Otoskeskiarvojen jakauma on suunnilleen normaalijakautunut, huolimatta muuttujan jakaumasta perusjoukossa (Saaranen 2018)

Keskiarvo Mean. Kuvaa havaintoarvojen suuruutta suunnilleen (Nummenmaa 2011)

Keskiarvon keskivirhe Standard error of mean. Tarkoittaa otoskeskiarvossa olevaa otantajakauman hajontaa (Saaranen 2018)

Keskiarvon luottamusväli Keskiarvon virhemarginaali, confidence interval. Yleistettäessä otoksen keskiarvo perusjoukkoon, tulee huomioida otantavirheen epävarmuus, jota mitataan virhemarginaalilla (Taanila 2019)

Keskihajonta Standard deviation. Kertoo havaintojen keskimääräisestä etäisyydestä jakauman keskiarvoon nähden (Nummenmaa 2011)

Kolmannen osapuolen data Ostetaan yrityksiltä, jotka eivät ole alun perin itse keränneet dataa esimerkiksi markkinatutkimusyritykset (Sponder & Khan 2018)

Korrelaatio Kahden tai useamman muuttujan välisen riippuvuuden voimakkuus (Saaranen 2018)

Korrelaatiokerroin Tulomomenttikorrelaatiokerroin, R , r . Tunnusluku muuttujien välisen riippuvuuden voimakkuudelle (Saaranen 2018)

Kuluttaja Henkilö, joka käyttää hyödykkeitä (Raijas 2016)

Kuluttajan ostopäätösprosessi Viisivaiheinen malli: Tarpeen tunnistaminen > Tiedon etsiminen > Vaihtoehtojen vertailu > Ostopäätös > Ostons jälkeinen käyttäytyminen (Kotler & Keller 2016)

Kuutiomalli Cubic model. Regressiomallin epälineaarinen kolmannen asteen käyrä (Saaranen 2017)

Kuvaileva tilastanalyysi Tunnusluvut, grafiikka, jakaumat ja riippuvuudet (Saaranen 2018)

Laatikko-jana-kuvio Boxplot. Kuviossa puolet havainnosta sisältyvät laatikkoon, johon mediaani on merkitty poikkiviivalla. Laatikon yläosa kuvaa yläkvartaalia ja alaosa alakvartaalia. Janan päiden välissä ovat muut havainnot, paitsi voimakkaimmin poikkeavat havainnot, jotka on merkitty ympyrällä (Nummenmaa 2011)

Lineaarinen regressio Suhteellisen yksikertainen tilastollinen menetelmä mallintaa lineaarinen yhteys yhden tai useamman selittävän muuttujan ja selitettävän muuttujan välillä (Sharda ym. 2018)

Long tail -hakusana muodostuu useammasta, yleensä kahdesta tai kolmesta erillisestä sanasta tai lauseesta (Rouhiainen 2019)

Lyhyen aikavälin ennustaminen Short-term forecasting. Tarkoittaa ennustamista 1–3 kuukauden päähän (Mahadevan (2010, teoksessa Kinski 2016, 16))

Mediaani Median. Kuvaa keskimmäistä havaintoa (Nummenmaa 2011)

Merkitsevyystaso Riskitaso, p-arvo, sig-arvo, sig-luku. Ilmaisee tuloksen tilastollisen merkitsevyyden tai yleistettäessä riippuvuus koko perusjoukkoon, virheen todennäköisyyden eli riskin sille, että havaittu riippuvuus tai ero johtuu sattumasta (Saaranen 2018)

Muuttuja Ominaisuus esimerkiksi numero, symboli tai määrä, jota voi mitata tai jonka voi laskea (Sharda ym. 2018)

Normaalisti jakautunut kuvaaja Kuvaajan jakauma on symmetrinen ja huippu osuu keskimmäiseen arvoon (Saaranen 2018)

Nykyisyyden ennustaminen Nowcasting. Tarkoittaa ennustamista enintään kolmen viikon päähän (Choin & Varian 2012)

Omni-channel Kaikkikanavaisuus

Outdoor Impact Ulkomainonnan yleisönmittausjärjestelmä (Clear Channel 2020)

Paneelidata Paneelissa mukana olevien henkilöiden tietokoneen käyttöä mittaavaa dataa (Kaushik 2007)

Pearsonin korrelaatiokerron Tunnusluku, joka mittaa numeeristen muuttujien lineaarisen riippuvuuden voimakkuutta (Saaranen 2018)

Persentiili Percentile eli sadannes ilmaistaan muodossa p %:n fraktiili, jolla tarkoitetaan muuttujan arvoa, jonka alle jää p % havainnoista; p on luku välillä 0–100 (Saaranen 2018)

POS-mainonta Point-Of-Sales. Ostopaikan lähelle sijoitettu mainonta (Riipinen 2019)

Pylväsdiagrammi Pylväskaavio, bar. Kuvaa eri luokkiin kuuluvia havaintoja, mitä korkeampi pylväs, sitä enemmän havaintoja (Nummenmaa 2011)

Päivittäistavarat Elintarvikkeet ja kulutustuotteista hygieniatuotteet, teknokemian- ja tupakkatuotteet sekä lehdet ja päivittäiskosmetiikka (Päivittäistavarakauppa 2020)

Regressioanalyysi Menetelmällä tutkitaan selittävän muuttujan tai selittävien muuttujien vaikutus selitettävään muuttujaan (Saaranen 2018)

Ristikorrelaatioanalyysi CCF, Cross-Correlation Analysis. Menetelmällä tutkitaan kahden aikasarjamuuttujan välistä riippuvuutta (Holmes, Scheuerell & Ward 2020)

Selitettävä muuttuja Dependent variable, y. Selitettävän muuttujan arvojen vaihtelua selitetään selittävällä muuttujalla (Nummenmaa 2011)

Selittävä muuttuja Independent Variable, x. Selittävällä muuttujalla selitetään selitettävän muuttujan arvojen vaihtelua (Nummenmaa 2011)

Selitysaste Selitysosuus, selityskerroin, R Square, R^2 , r^2 . Kertoo selittävän muuttujan kyvyn selittää selitettävän muuttujan vaihtelua (Saaranen 2018)

Semistrukturoitu data Strukturoidun ja strukturoimattoman datan välimuoto (Marr 2016)

Sisäinen data Yrityksen sisäistä dataa asiakkaista ja yrityksen toiminnoista

Spatiotemporaalinen analyysi Uuden tiedon tuottaminen käsittelemällä tiettyyn paikkaan tai alueeseen viittaavaa tietoa ja tähän sisältyvää aikatietoa (Sanastokeskus TSK 2020)

Strukturoimaton data Tekstiä, valokuvia, puhetta tai videoita (Taanila 2019)

Strukturoitu data Määrällistä tai kategorista dataa, jonka voi tallentaa tietokantaan tai taulukkoon (Taanila 2019)

Tilastollinen analyysi Kuvaileva analyysi ja tilastollinen päättely (Saaranen 2018)

Tilastollinen päättely Kuinka todennäköisesti tulokset voidaan yleistää perusjoukkoon (Saaranen 2018)

Toisen osapuolen data Kerätään myyntiin suoraan muilta organisaatioilta, yrityksiltä ja yksilöiltä (Sponder & Khan 2018)

Transaktiodata Yrityksen liiketoimintaprosessista, esimerkiksi ostoista ja tilauksista, kertyvä data (Finanssiala 2019)

Ubersuggest Avoin hakukoneanalytiikkatyökalu muun muassa hakusanatutkimuksen ja hakuvolyymien kartoittamiseen (Ubersuggest 2019)

Ulkoinen data Yrityksen ulkopuolista dataa kilpailijoista, markkinoista ja kuluttajista

Ulkomainonta OOH, Out-Of-Home. Kaikki kodin ulkopuolinen media, kuten esimerkiksi kauppakeskuksissa, rakennusten ulkoseinissä ja kadun varsilla sijaitsevat mainospinnat (Mediateko 2020)

VAC Visibility Adjusted Contact. Mediavaluutta ilmaisee todennetut katsekontaktit eri ulkomedioissa ja mediayhdistelmissä havainnointitutkimuksen mallinnusmenetelmän mukaan (Outdoor Impact 2020)

Vakiotekijä Constant

Verkkoanalytiikka Verkkosivujen trafiikin ja sivuston konversion mittaamista ja analysointia (Sponder & Khan 2018)

Vertikaali Kapean sektorin media (Sulopuisto 2018)

Viive Lag. Aika Google-hausta ostoon

Viiveluku Lag Number. Mittaa aikaa Google-hausta oston

Vinousluku Skewness. Kuvaa normaalisti jakautuvan kuvaajan poikkeamaa: positiivinen arvo vinous oikealle ja negatiivinen arvo vinous vasemmalle (Saaranen 2018)

Ärsyke Herätteen seurauksena kuluttaja havaitsee tarpeen ja motivoituu toimimaan (Bergström & Leppänen 2015)

Liite 2. Sesonkituotteiden kartoitus: hakumäärät ja -piikit

Lähteet: Tilastokeskus 2019d, 2019e, 2019f & 2018; Ubersuggest 28.11.2019

Sesonki	Sesonkituote	Hakumäärät/ -volyyymi	Hakupiikki/-piikit
Pääsiäinen 19.–22.4.2019	mämmi	6600	33100 huhtikuu
	lammas	3600	5400 huhtikuu
	pasha	0	
	kananmunat	4400	5400 tammi-, maalis- ja huhtikuu
	rahka	2400	2900 tammi-, maalis- ja huhtikuu
Vappu 1.5.2019	sima	8100	74000 huhtikuu
	tippaleivät	1900	14800 huhtikuu
	munkit	5400	33100 huhtikuu
	kuohuviini	2400	3600 joului- ja toukokuu
	samppanja	1900	2900 joulukuu 2400 toukokuu
	perunasalaatti	12100	40500 joulukuu 27100 huhtikuu
	nakki	1900	2400 huhtikuu
Joulu 24.–26.12.2018	joulukinkku	4400	40500 joulukuu
	kinkkurulla	1000	4400 joulukuu
	joulutorttu	8100	60500 joulukuu
	joulupipari (pipari)	720 (1900)	6600 (12100) joulukuu
	lanttulaatikko	12100	135000 joulukuu
	porkkanalaatikko	18100	165000 joulukuu
	perunalaatikko	8100	49500 joulukuu
	suklaakonvehdit	590	2900 joulukuu
	olut	3600	3600 joului-, maalis-, touko-, kesä- ja heinäkuu
	lipeäkala	2900	18100 joulukuu
	punaviini	1600	2900 joulukuu
Uusi sato kevät–kesä 2019	mansikat	6600	18100 heinäkuu
	varhaisperuna	720	4400 kesäkuu
	varhaiskaali	1000	4400 kesäkuu
	tankoparsa	1900	8100 toukokuu
	valkoinen parsa	390	1900 huhtikuu
	vihreä parsa	390	1900 huhtikuu
	retiisi	1900	2900 kesä- ja heinäkuu
	raparperi	5400	22200 kesäkuu
	herne	1900	4400 heinäkuu

Liite 3. Datojen mallitaulukot

Esimerkki Google Trendistä kerätystä datasta csv-muodossa Excelissä

	A	B	C	D	E	F	G
1	Arvot: Ruoka ja juoma						
2							
3	Aika,Nyhtökaura: (Suomi),Soijarouhe: (Suomi),Härkis: (Suomi)						
4	2020-01-15T14,23,15,26						
5	2020-01-15T15,20,45,45						
6	2020-01-15T16,41,23,43						
7	2020-01-15T17,59,48,32						
8	2020-01-15T18,33,32,24						

Google Trendistä kerättyä esimerkkitdataa Excelissä, jossa on lisätty funktiotoiminnolla kuukaudet ja viikot

	A	B	C	D	E	F	G
1	Pvm 19	Kk 19	Vk 19	Indeksi 19	Pvm 18	Kk 18	Vk 18
2	2019-01-06	1	1	100	2018-01-07	1	1
3	2019-01-13	1	2	58	2018-01-14	1	2
4	2019-01-20	1	3	67	2018-01-21	1	3
5	2019-01-27	1	4	56	2018-01-28	1	4
6	2019-02-03	2	5	54	2018-02-04	2	5
7	2019-02-10	2	6	53	2018-02-11	2	6
8	2019-02-17	2	7	76	2018-02-18	2	7

Google Trendistä kerättyä esimerkkitdataa Excelissä viikonpäivien keskiarvojen selvittämiseksi

	A	B	C	D	E	F	G	H	I
1	Pvm voi	Päivä voi	Indeksi voi	Pvm JT	Päivä JT	Indeksi JT	Pvm JST	Päivä JST	Indeksi JST
2	2019-11-18	2	33	2019-11-18	2	41	2019-11-18	2	38
3	2019-11-19	3	84	2019-11-19	3	41	2019-11-19	3	26
4	2019-11-20	4	43	2019-11-20	4	63	2019-11-20	4	48
5	2019-11-21	5	39	2019-11-21	5	79	2019-11-21	5	32
6	2019-11-22	6	27	2019-11-22	6	59	2019-11-22	6	28

Google Trendistä kerättyä ei-sesonkituotteiden esimerkkitdataa Excelissä kellonaikojen keskiarvojen selvittämiseksi seitsemän päivän ajanjaksolla

	A	B	C
1	Pvm	Aika	Indeksi
2	2020-01-15	12	20
3	2020-01-15	13	24
4	2020-01-15	14	36
5	2020-01-15	15	37
6	2020-01-15	16	43

Google Trendistä kerättyä sesonkituotteiden esimerkkidataa Excelissä eri päivien kellon-aikojen selvittämiseksi hakuhuippuun sijoittuvana seitsemän vuorokauden ajanjaksona

	A	B	C	D	E	F
1	Pvm JK	Indeksi JK	Pvm LL	Indeksi LL	Pvm PL	Indeksi PL
2	2019-12-18T17	5	2019-12-18T17	30	2019-12-18T18	25
3	2019-12-18T18	5	2019-12-18T18	28	2019-12-18T19	22
4	2019-12-18T19	6	2019-12-18T19	26	2019-12-18T20	18
5	2019-12-18T20	6	2019-12-18T20	21	2019-12-18T21	16

Osa Alkon myyntidatojen ja Google Trends -indeksien Excel-kokoomataulukkoa

	A	Q	R	S	T	U	V	W	X	Y	Z
1	Aika	Myynti Po	Myynti Ke	Myynti Pohj	Myynti Kai	Myynti Lap	Indeksi Su	Indeksi Uu	Indeksi Var	Indeksi Sa	Indeksi Ka
2	2019-10	56000	18000	87000	19000	56000	27	20	18	22	0
3	2019-10	56000	18000	87000	19000	56000	29	27	18	0	0
4	2019-10	56000	18000	87000	19000	56000	30	30	23	0	0
5	2019-10	56000	18000	87000	19000	56000	28	52	28	0	0
6	2019-09	48000	15000	78000	18000	65000	23	40	27	0	28

Liite 4. Ohjelmien ja palveluiden asetukset

Tunnuslukuyhteenvedot punaviinin Google Trends -indeksi- ja myyntidatasta Suomessa
SPSS: Analyze > Descriptive Statistics > Frequencies > Variables: Indeksii Suomi, Myynti Suomi > Frequencies: Statistics: Percentile Values: Quartiles, Percentiles: 10, 25, 50, 75, 90 > Add, Central Tendency: Mean, Median, Mode, Dispersion: Std. Deviation, Variance, Range, Minimum, Maximum > Continue > OK (SPSS 2020).

Tuotenimikkeen hakemisen kuukausikohtaiset erot viiden vuoden Google Trends -hakuindeksien keskiarvoista laatikko-janakuvioilla

SPSS > Graphs > Legacy Dialogs > Boxplot > Simple, Summaries for groups of cases > Other Statistics, Variable: Indeksii tuote, Category Axis: Kk (kuukausi) (SPSS 2020).

Tuotenimikkeen hakemisen viikkojen väliset erot Google Trends -hakuindeksien viiden vuoden keskiarvoista pylväsdiagrammeina

SPSS > Graphs > Legacy Dialogs > Bar Charts: Simple, Summaries for groups of cases > Bars Represent: Other Statistics, Variable: Indeksii tuote, Category Axis: Vk (viikko) (SPSS 2020).

Tuotenimikkeen hakemisen viikonpäivien väliset erot Google Trends -hakuindeksien 90 päivän keskiarvoista pylväsdiagrammeina

SPSS > Graphs > Legacy Dialogs > Bar Charts > Simple > Summaries of group of cases > Bars Represent/Other statistics, Variable: Indeksii tuote, Category Axis: Päivä tuote (SPSS 2020).

Ei-sesonkituotteiden hakemisen kellonaikojen väliset erot Google Trends -hakuindeksien seitsemän päivän keskiarvoista pylväsdiagrammeina

SPSS > Graphs > Legacy Dialogs > Bar Charts > Simple > Summaries of group of cases > Bars Represent/Other statistics, Variable: Indeksii tuote, Category Axis: Aika tuote (SPSS 2020).

Sesonkituotteiden hakemisen päivien ja kellonaikojen väliset erot Google Trends -hakuindekseistä seitsemänä päivänä hakuhiipun aikaan pylväsdiagrammeina

SPSS > Graphs > Legacy Dialogs > Bar Charts > Simple > Summaries for group of cases > Bars Represent/Other statistics, Variable: Indeksii tuote, Category Axis: Pvm tuote (SPSS 2020).

Tuotenimikkeiden kuukausien, viikkonumeroiden, viikonpäivien ja kellonaikojen keskiarvoon perustuvien Google Trends -indeksien tulosten luottamusvälit

SPSS: Legacy Dialogs > Error Bar > Simple, Summaries for Groups of Cases > Define > Simple Error Bar: Variable: Indeksi tuote, Category Axis: kk, vk, pv tai aika tuote, Bars Represent: Level 95 % (Luottamustasoksi valittiin 95 %) > OK (SPSS 2020; Saaranen 2017, 64–65).

Punaviinin myynnin ja hakemisen ristikorrelaatioanalyysi Suomessa

SPSS: Analyze > Forecasting > Cross-correlation: Variables: Myynti Suomi ja Indeksi Suomi > OK (SPSS 2020).

Punaviinin myynnin ja hakemisen ristikorrelaatioanalyysi maakunnissa

SPSS > Analyze > Forecasting > Cross-Correlations: Variables: Myynti Maakunta ja Indeksi Maakunta > OK (SPSS 2020).

Maakuntien punaviinin Google-hakujen ja myyntien välisiä viive-eroja havainnollistavat kartat

Tilastokeskuksen Tee oma karttaesitys > Valittiin karttapohja: Maakunnat 2017 > Aineistot > Käsittele aineistoja > Lataa Excel-tiedosto > Lataa > Numeeriset muuttujat: valitaan Viive 0, Viive 1 ja Viive 2 > Lataa. Palvelun etusivulle aukeaa Viive 0-kartta, ja kaksi muuta karttaa saa näkyviin valikosta. (Tilastokeskus 2020a.)

Kahden muuttujan välisen riippuvuuden tutkiminen hajontakuviolla

SPSS: Graphs > Legacy Dialogs > Scatter/Dot > Simple Scatter > Define > Simple Scatterplot: Y Axis: Myynti Suomi, X Axis: Indeksi Suomi > OK (SPSS 2020).

Korrelaatiomatriisin toteuttaminen

SPSS: Analyze > Correlate > Bivariate > Bivariate Correlations: Variables: Indeksi Suomi + maakunnat ja Myynti Suomi + maakunnat, Correlation Coefficients: Pearson, Test of Significance: Two-tailed > OK (SPSS 2020; Saaranen 2017, 55).

Lineaarisen regressioanalyysin toteuttaminen

SPSS: Analyze > Regression > Linear > Linear Regression: Dependent: Myynti Suomi, Independent: Indeksi Suomi, Plots: Histogram, Normal probability plot > Continue > OK (SPSS 2020; Saaranen 2017, 58).

Hajontakuvion muodostaminen

SPSS: Graps > Legacy Dialogs > Scatter/Dot > Simple > Define > Simple Scatterplot: Y Axis: Myynti Suomi, X Axis: Indeksii Suomi > OK. Lisättiin suora kuvioon: kuvion tuplaklikkaus > hiiren oikealla olevan painikkeen painaminen pisteparven kohdalla > Add Fit Line at Total > Properties-ikkunan sulkeminen. (SPSS 2020; Saaranen 2017, 57.)

Ennusteen keskiarvojen luottamusvälit havaintojen perusteella

SPSS: Analyze > Regression > Linear > Linear Regression: Dependent: Myynti Suomi, Independent: Indeksii Suomi > Save: Predicted Values: Unstandardized, Prediction Intervals: Mean > OK (SPSS 2020; Taanila 2010, 17–18).

Mallin ennusteiden tarkkuuden arvioiminen havainnoista lasketuista ennusteiden luottamusväleistä

SPSS: Analyze > Regression > Linear > Linear Regression: Dependent: Myynti Suomi, Independent: Indeksii Suomi > Save > Linear Regression: Save: Predicted Values: Unstandardized, Residuals: Unstandardized, Prediction Intervals: Mean > Continue > OK (SPSS 2020; Taanila 2010, 17).

Tunnuslukuyhteenvedot virhetermeille

SPSS: Analyze > Descriptive Statistics > Frequencies > Variables: RES_1 > Frequencies: Statistics: Percentile Values: Quartiles, Percentiles: 25, 50, 75 > Add, Central Tendency: Median, Mode, Dispersion: Std. Deviation, Minimum, Maximum > Continue > OK (SPSS 2020; Saaranen 18.3.2020).

Jäännöstermien jakauma

SPSS: Analyze > Regression > Linear > Linear Regression: Dependent: Myynti Suomi, Independent: Indeksii Suomi, Plots > Linear Regression: Plots: Standardized Residual Plots: Valitaan Histogram ja Normal probability plot > Continue > OK (SPSS 2020).

Ominaisuuksien tilastollisen merkitsevyyden testaaminen

SPSS: Analyze > Descriptive Statistics > Descriptives: Variables: Indeksii Suomi, Myynti Suomi > Descriptive: Options: Characterize Posterior Distribution: Valitaan Kurtosis, Skewness. Display Order: Variable list > Continue > OK (SPSS 2020).

Lineaarinen riippuvuus ja jäännösten varianssien yhtäsuuruus jäännöskuviolla

SPSS: Analyze > Regression > Linear > Linear Regression: Dependent: Myynti Suomi, Independent: Indeksii Suomi, Plots > Y: *ZRESID, X: *ZPRED > Continue > OK (SPSS 2020; Taanila 2010, 12).

Jäännösten normaalijakautuneisuuden tarkastelu kuviolla

SPSS: Analyze > Regression > Linear > Linear Regression: Dependent: Myynti Suomi, Independent: Indeksii Suomi, Plots > Y: *ZRESID, X: *ZPRED, Standardized Residual Plots: Normal probability plot Continue > OK (SPSS 2020; Taanila 2010, 14).

Jäännösten riippumattomuuden testaus Durbin-Watsonin testillä

SPSS: Analyze > Regression > Linear > Linear Regression: Independent Variable Indeksii Suomi, Dependent Variable Myynti Suomi > Statistics > Linear Regression Statistics: Regression Coefficients: Estimates, Model fit, Residuals: Durbin-Watson > OK (SPSS 2020; Karadimitriou & Marshall 2020).

Epälineaarisen käyräsovituksen toteuttaminen

SPSS: Analyze > Regression > Curve Estimation: Dependent: Myynti Suomi, Variable: Indeksii Suomi, Models: valittiin kaikki mallivaihtoehdot > OK (SPSS 2020; Saaranen 2017, 61).

Kuutiomallia kuvaavan käyrän muodostaminen

SPSS: Analyze > Regression > Curve Estimation: Dependent: Myynti Suomi, Variable: Indeksii Suomi, Models: Cubic (poistetaan rastit muiden mallien kohdalta) > OK (SPSS 2020).

Liite 5. Elintarvikeryhmien tuotenimikkeiden sekä hakusanojen ja -volyyymien määrittäminen

Meijerituotteiden hakusanat COICOP-luokituksen mukaan, hakumäärät ja valitut Google Trends -hakusanat

Lähteet: Tilastokeskus 2019b; Übersuggest 28.11.2019

COICOP-luokitus	Hakusana	Hakumäärät/ volyyymi	Valitut hakusanat Google Trends
01.1.4 Maitotuotteet, juusto ja kananmunat	maitotuotteet maitotuote meijerituotteet meijerituote	390 390 20 0	
01.1.4.1 Maito	maito	1900	maito
01.1.4.1.2 Vähärasvainen maito	vähärasvainen maito	0	
01.1.4.1.2.100 Kevytmaito	kevytmaito	590	
01.1.4.1.2.200 Rasvaton maito	rasvaton maito	720	
01.1.4.3 Jogurtti	jogurtti jugurtti	1000 720	jogurtti jugurtti
01.1.4.3.1 Jogurtti			
01.1.4.3.1.100 Jogurttipurkki	jogurttipurkki	50	
01.1.4.3.1.200 Jogurttitölkki	jogurttitölkki	0	
01.1.4.4 Juustot ja rahkat	juustot juusto rahkat rahka	2400 2400 2400 2400	juusto rahka
01.1.4.4.1 Kypsytytetyt juustot	kypsytytetyt juustot kypsytytty juusto	70 110	
01.1.4.4.1.100 Edamjuusto	edamjuusto	320	
01.1.4.4.1.200 Emmentalijuusto	emmentalijuusto	20	
01.1.4.4.1.300 Kermajuusto	kermajuusto	390	
01.1.4.4.1.400 Homejuusto	homejuusto	590	
01.1.4.4.1.500 Salaattijuusto	salaattijuusto	0	
01.1.4.4.2 Tuorejuustot ja rahkat			
01.1.4.4.2.100 Tuorejuusto	tuorejuusto	1900	
01.1.4.4.2.200 Raejuusto	raejuusto	1600	
01.1.4.4.2.300 Maitorahka	maitorahka	1300	
01.1.4.4.3 Sulatejuustot	sulatejuustot	20	
01.1.4.4.3.100 Sulatejuusto	sulatejuusto	880	
01.1.4.5 Muut maitotaloustuotteet	maitotaloustuotteet maitotaloustuote	20 0	
01.1.4.5.1 Kermat ja kermavalmisteet	kermat kermavalmisteet kerma kermavalmiste	70 0 1300 0	kerma
01.1.4.5.1.100 Kuohukerma	kuohukerma vispikerma	1300 880	
01.1.4.5.1.200 Ruokakerma	ruokakerma	880	
01.1.4.5.1.300 Hapankermavalmiste	hapankerma kermaviili smetana ranskankerma	480 1300 2900 1900	hapankerma kermaviili smetana ranskankerma
01.1.4.5.2 Maitopohjaiset jälkiruoat	maitopohjaiset jälkiruoat	0 0	

	maitopohjainen jälki-ruoka		
01.1.4.5.2.100 Jälkiruokavanukas	jälkiruokavanukas maitovanukas kermavanukas vanukas	10 110 110 720	
01.1.4.5.2.200 Jälkiruokarahka	jälkiruokarahka	170	
01.1.4.5.3 Hapanmaitotuotteet	hapanmaitotuotteet hapanmaitotuote	480 480	hapanmaitotuote
01.1.4.5.3.100 Piimä	piimä	1600	piimä
01.1.4.5.3.200 Viili	viili	720	viili
01.1.4.5.4 Maitopohjaiset juomat	maitopohjaiset juomat maitopohjainen juoma	0 0	
01.1.4.5.4.100 Laktoositon maitojuoma	laktoositon maitojuoma laktoositon maito	70 480	laktoositon maito
01.1.5 Öljyt ja rasvat			
01.1.5.1 Voi	voi	5400	voi
01.1.5.1.1 Meijerivoi	meijerivoi	70	
01.1.5.1.1.100 Meijerivoi			
01.1.8 Sokeri, hillot, hunaja, suklaa ja makeiset			
01.1.8.5 Jäätelöt ja sorbetit			
01.1.8.5.1 Jäätelöt	jäätelöt jäätelö kermajäätelö maitojäätelö	4400 4400 90 90	jäätelö kermajäätelö maitojäätelö
01.1.8.5.1.100 Jäätelöpaketti	jäätelöpaketti	90	
01.1.8.5.1.200 Gourmet-jäätelö	gourmet-jäätelö herkuttelujäätelö	0 0	
01.1.8.5.1.300 Jäätelötuutti	jäätelötuutti	210	
01.1.8.5.1.400 Jäätelöpuikko	jäätelöpuikko	110	

Valmisaterioiden hakusanat COICOP-luokituksen mukaan, hakusanat ja -määrät ja valitut hakusanat Google Trends -hakuja varten

Lähteet: Tilastokeskus 2019c; Übersuggest 28.11.2019

ATERIAT	Hakusanat ja hakumäärät/-volyymit		Valitut Google Trends -hakusanat
Laatikot, keitot, salaatit, pastaruokat			
	laatikkoruoka laatikkoruokat laatikkoruuat	1900 1900 2400	laatikkoruuat
M0111102 Maksalaatikko	maksalaatikko maksalaatikko Saarioinen maksalaatikko Atria maksalaatikko Pirkka maksalaatikko Rainbow maksalaatikko Snellman	5400 210 20 10 0 0	maksalaatikko
M0111103 Riisivalmisteet	riisivalmiste	0	riisivalmiste
M0111303 Lihamakaroni- laatikko ym. pastaruokat	lihamakaronilaatikko Saarioinen Atria	4400 50 20	lihamakaronilaatikko, pastaruoka, pasta-ate- ria

	Pirkka Kotikylä Kokkikartano pastaruoka pastaruokat pastaruuat	20 0 0 1000 1000 1000	
M0112704 Lihakaali-, li- haperunase- ym. laati- kot	lihakaalilaatikko Saarioinen Kokkikartano kaalilaatikko Pirkka HK lihaperunaseleaatikko Atria Saarioinen HK kinkkukiusaus Saarioinen Kokkikartano HK Pirkka	320 - - 33100 30 10 4400 - - - 22200 - - 0 70	lihakaalilaatikko, liha- perunaseleaatikko, kinkkukiusaus
M0112707 Eines- ja pa- kastekeitot lihasta	lihakeitto hernekeitto Jalostaja Rainbow Euroshopper Pirkka	6600 12100 90 0 0 10	lihakeitto
M0112708 Kana-, kinkku- ym. salaattit lihasta	kanasalaatti kinkkusalaatti	6600 720	kanasalaatti, kinkku- salaatti
M0112710 Valmiit ruoka- annokset lihasta			
M0113405 Silakkalaa- tikko, janssonin kiusaus ym.	silakkalaatikko silakkalaatikko Saarioinen silakkalaatikko Pirkka janssonin kiusaus Kokkikartano janssoninkiusaus HK lohikiusaus Kokkikartano HK Pirkka	2400 30 10 8100 0 9900 0 4400 - 0 30	silakkalaatikko, jans- sonin kiusaus
M0113406 Kala-, katka- rapu- ym. salaattit	kalasalaatti katkarapusalaatti	170 2900	kalasalaatti, katkara- pusalaatti
M0113407 Valmiit ruoka- annokset kalasta			
M0113408 Kalakeitot ja muut kalaeinekset	kalakeitto Apetit lohikeitto Kokkikartano Pirkka Saarioinen	12100 30 14800 30 10 10	kalakeitto
M0117605 Valmiit ruoka- annokset kasviksista			
M0117607 Kasviskeitot, -laatikot ja -einekset	kasviskeitto kasvisselekeitto Atria Rainbow Pirkka Apetit bataattiselekeitto Saarioinen	2400 14800 0 10 20 30 30 12100	kasvisselekeitto lanttulaatikko porkkanalaatikko perunaseleaatikko

	pinaattikeitto	140	
	pinaattikeitto pakaste	110	
	Apetit	40	
	Pirkka	-	
	Saarioinen	-	
	Kokkikartano	720	
	kasvishernekeitto	10	
	Jalostaja	10	
	Kymppi	480	
	kasvislaatikko	12100	
	lanttulaatikko	90	
	Saarioinen	10	
	Atria	10	
	Kymppi	18100	
	porkkanalaatikko	10	
	Kymppi	10	
	Atria	110	
	Saarioinen	8100	
	perunalaatikko	30	
	Saarioinen	10	
	Kymppi	0	
	Lihajaloste Korpela	0	
	Korpela	12100	
	imelletty perunalaatikko	50	
	Saarioinen	0	
	(Lihajaloste) Korpela	5400	
	punajuurilaatikko	10	
	Kymppi	14800	
	pinaattiletut	140	
	Pirkka	390	
	pinaattiohukainen	0	
	Atria	0	
	Pirkka	390	
	pinaattiohukaiset	0	
	Kruunu Herkku		

Kasviproteiinien luokittelu, hakusanat ja -määrät sekä valitut hakusanat Google Trends -hakuihin

Lähteet: vegaanituotteet.net 2019; Ubersuggest 28.11.2019

Luokittelu	Hakusanat ja hakumäärät/-volyymit		Valitut Google Trends -hakusanat
Yleistermit	kasviproteiini(t)	390	kasvisproteiini lihankorvike
	kasviproteiinivalmisteet	40	
	kasviproteiinituotteet	10	
	kasviproteiinijaloste	0	
	lihankorvike	260	
	lihan korvike	170	
	maidonkorvike	90	
	maidon korvike	90	
Lihan sijaan käytettävät	Nyhtökaura	6600	Nyhtökaura Härkis Härkispihvit Härkispörykät seitan kinkku seitan kebab
	Härkis	3600	
	pihvit	590	
	pyörykät	320	
	Nyhtis	110	
	seitan	0	

	seitan kinkku seitan kebab seitanpaisti seitan pihvi seitanmakkara Kaurajauhis mantelijauho quorn tempe soija rouhe soijarouhe soijanakki soijanakko	1900 480 170 70 30 1300 1900 4400 0 3600 2900 320 10	Kaurajauhis mantelijauho quorn soija rouhe soijarouhe
Maidon sijaan käytettävät	kaurajuoma soijajuoma riisijuoma kookosjuoma kookosjogurtti soijakerma kaurakerma soijajogurtti soijarahka kauravälipala	880 110 70 90 390 210 1000 590 140 210	kaurakerma kaurajuoma soijajogurtti kookosjogurtti
Kasvispohjaiset juustot ja jäätelöt	vege juusto vegejuustot kasvisjuusto kasvis juusto maidoton juusto maidoton jäätelö kasvisjäätelö	100 170 50 50 480 480 20	maidoton juusto maidoton jäätelö

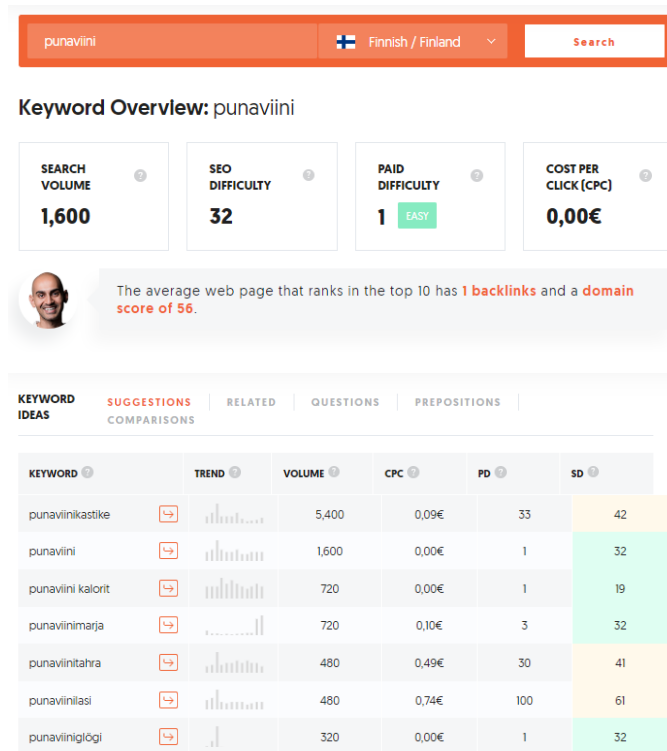
Joulusesongin tärkeimmät tuotteet, hakumäärät ja -piikit sekä yleisimmät hakuajankohdat

Lähteet: Tilastokeskus 2018a; Übersuggest 28.11.2019

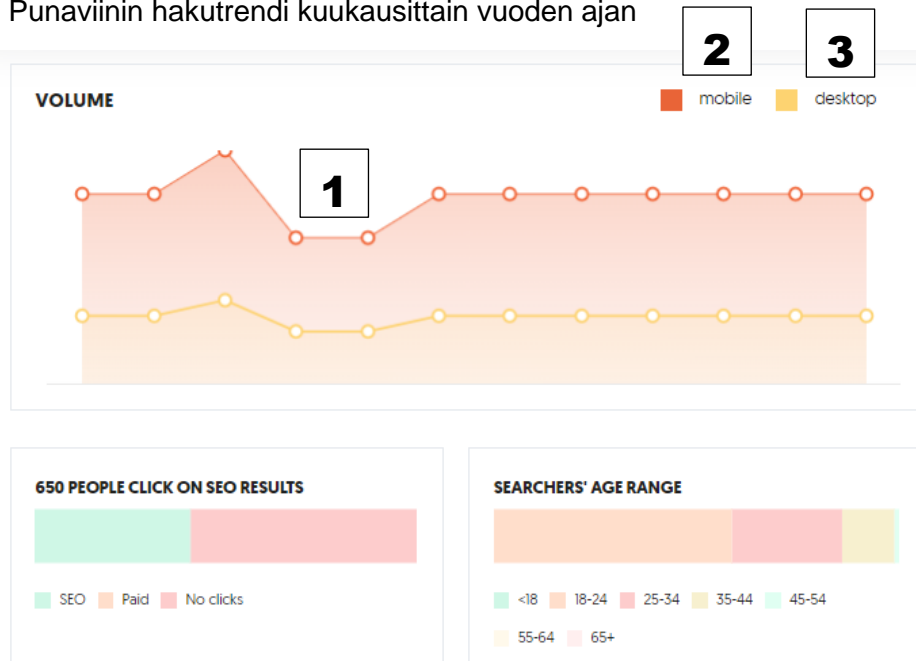
Tuotteet/hakusanat	Hakumäärät/ -volyymit	Hakuhuippu	Yleisin hakuajan- kohta
joulukinkku	4400	40500	joulukuu
kinkkurulla	1000	4400	joulukuu
joulutorttu	8100	60500	joulukuu
joulupipari	720	6600	joulukuu
pipari	1900	12100	joulukuu
lanttulaatikko	12100	135000	joulukuu
porkkanalaatikko	18100	165000	joulukuu
perunalaatikko	8100	49500	joulukuu
suklaakonvehdit	590	2900	joulukuu
olut	3600	3600	joulu-, maalis-, touko-, kesä- ja heinäkuu
lipeäkala	2900	18100	joulukuu
punaviini	1600	2900	joulukuu

Liite 6. Ubersuggest-työkalun tuottamat hakuvolyymit ja hakusanaideat

Lähde: Ubersuggest 15.1.2020



Punaviinin hakutrendi kuukausittain vuoden ajan



1	Kk	1	2	3	4	5	6	7	8	9	10	11	12
	Kpl	1300	1300	1600	1000	1000	1300	1300	1300	1300	1300	1300	1300
2	Mobiili	836	836	1029	643	643	836	836	836	836	836	836	836
3	Pöytä-kone	464	464	571	357	357	464	464	464	464	464	464	464

Liite 7. Tuotenimikkeiden hakijoiden ikäryhmät

Lähde: Ubersuggest 25.1.2020

	Ikäluokka						
	< 18	18–24	25–34	35–44	45–54	55–64	> 65
Tuotenimike	Osuus, %						
Voi	0	30,6	23,6	37,5	8,3	0	0
Jäätelö	0	40	30	30	0	0	0
Juusto	0	38,9	18,5	40,7	1,9	0	0
Maksalaatikko	0	18,9	29,5	48,7	0,6	2,3	0
Kinkkukiusaus	0	30,7	32,7	32,7	3,9	0	0
Kasvissosekeitto	0	18,4	47,7	31,6	2	0	0
Nyhtökaura	0	28,3	31,9	34,1	5,8	0	0
Soijarouhe	0	43,2	27,9	23,4	2,7	2,7	0
Härkis	0	23,2	28,4	46,3	2,1	0	0
Porkkanalaatikko	0	15,4	38,6	39,6	3,9	2,4	0
Lanttulaatikko	0	13	43,2	35,8	5,5	2,6	0
Joulukinkku	0	35,4	39,1	20,5	4	1	0
Punaviini	0	58,6	27,1	12,9	1,4	0	0

Liite 8. Maakuntien ristikorrelaatiotaulukot

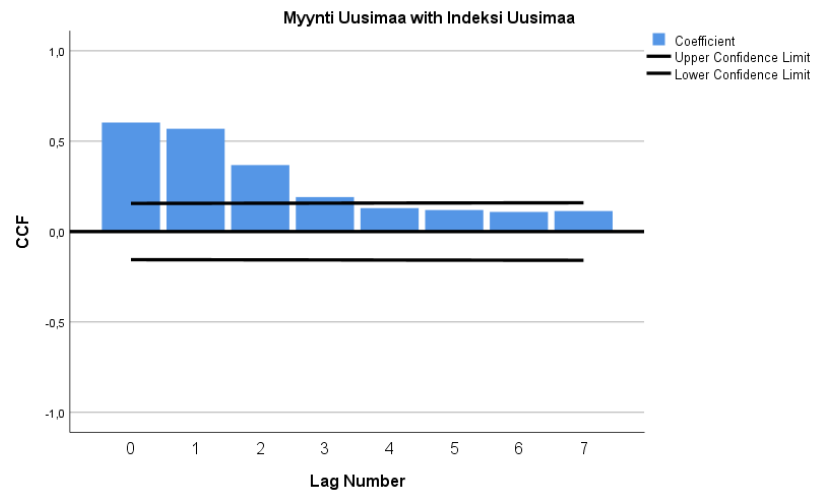
SPSS 2020

Cross Correlations

Series Pair: Myynti Uusimaa with Indeks Uusimaa

Lag	Cross Correlation	Std. Error ^a
-7	,000	,080
-6	,021	,079
-5	,037	,079
-4	,140	,079
-3	,279	,079
-2	,505	,078
-1	,574	,078
0	,603	,078
1	,569	,078
2	,368	,078
3	,191	,079
4	,130	,079
5	,119	,079
6	,109	,079
7	,114	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

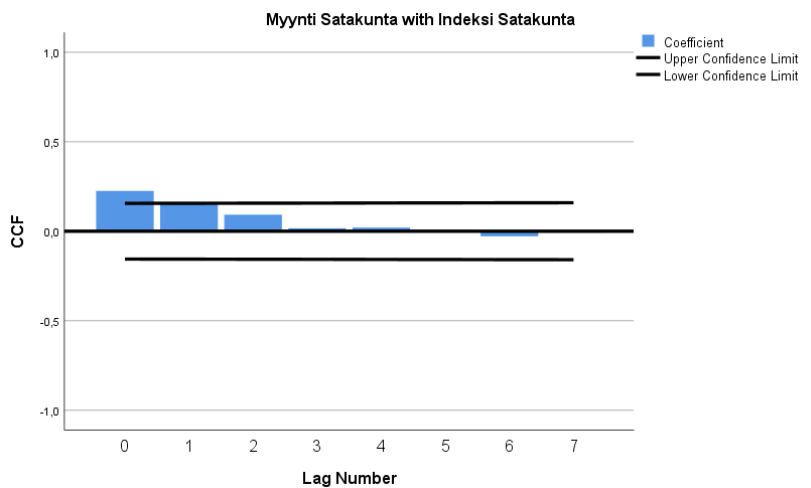


Cross Correlations

Series Pair: Myynti Satakunta with Indeks Satakunta

Lag	Cross Correlation	Std. Error ^a
-7	,082	,080
-6	,053	,079
-5	,053	,079
-4	,042	,079
-3	,015	,079
-2	,176	,078
-1	,225	,078
0	,225	,078
1	,160	,078
2	,092	,078
3	,016	,079
4	,020	,079
5	,005	,079
6	-,028	,079
7	-,007	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

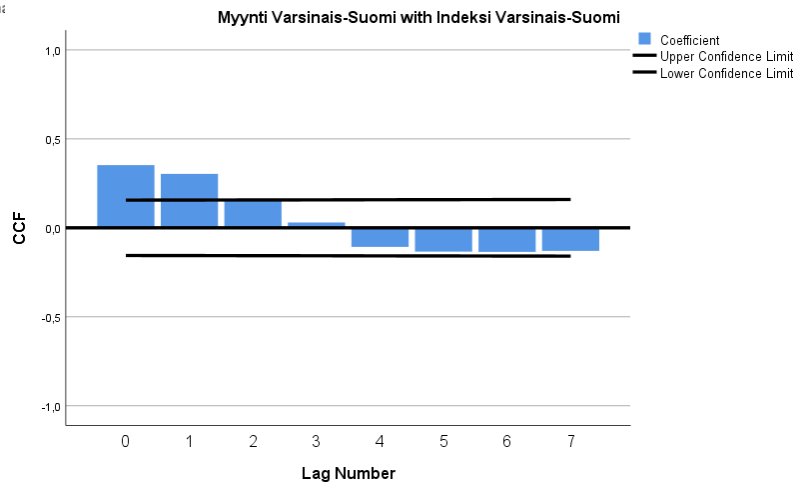


Cross Correlations

Series Pair: Myynti Varsinais-Suomi with Indeks Varsinais-Suomi

Lag	Cross Correlation	Std. Error ^a
-7	,018	,080
-6	,040	,079
-5	,071	,079
-4	,085	,079
-3	,196	,079
-2	,247	,078
-1	,300	,078
0	,352	,078
1	,303	,078
2	,152	,078
3	,030	,079
4	-,107	,079
5	-,134	,079
6	-,135	,079
7	-,130	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

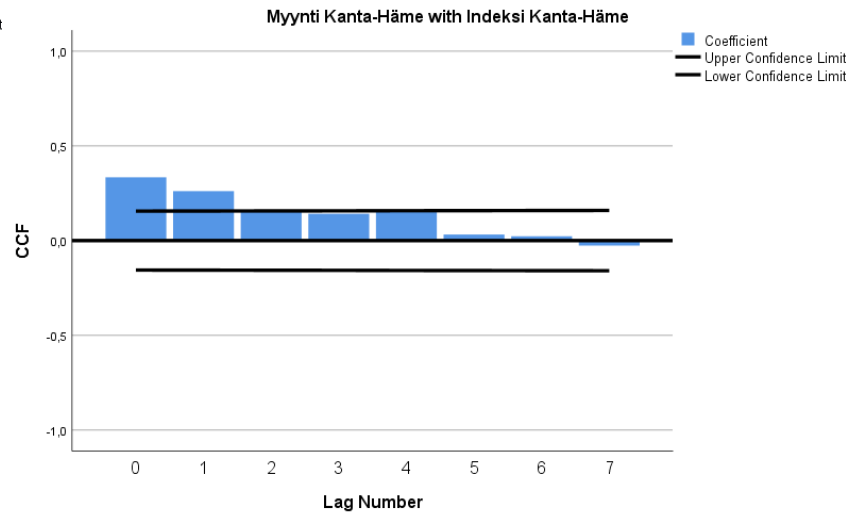


Cross Correlations

Series Pair: Myynti Kanta-Häme with Indeks Kanta-Häme

Lag	Cross Correlation	Std. Error ^a
-7	-,105	,080
-6	-,035	,079
-5	-,035	,079
-4	-,011	,079
-3	,073	,079
-2	,173	,078
-1	,165	,078
0	,334	,078
1	,261	,078
2	,155	,078
3	,141	,079
4	,161	,079
5	,032	,079
6	,023	,079
7	-,026	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

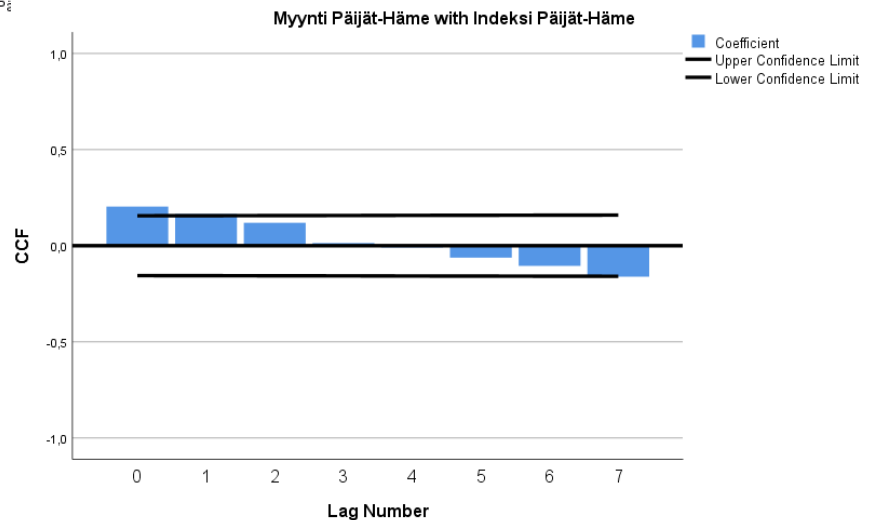


Cross Correlations

Series Pair: Myynti Päijät-Häme with Indeks Päijät-Häme

Lag	Cross Correlation	Std. Error ^a
-7	-,068	,080
-6	-,021	,079
-5	,039	,079
-4	,110	,079
-3	,168	,079
-2	,268	,078
-1	,190	,078
0	,203	,078
1	,165	,078
2	,119	,078
3	,014	,079
4	-,009	,079
5	-,062	,079
6	-,105	,079
7	-,161	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

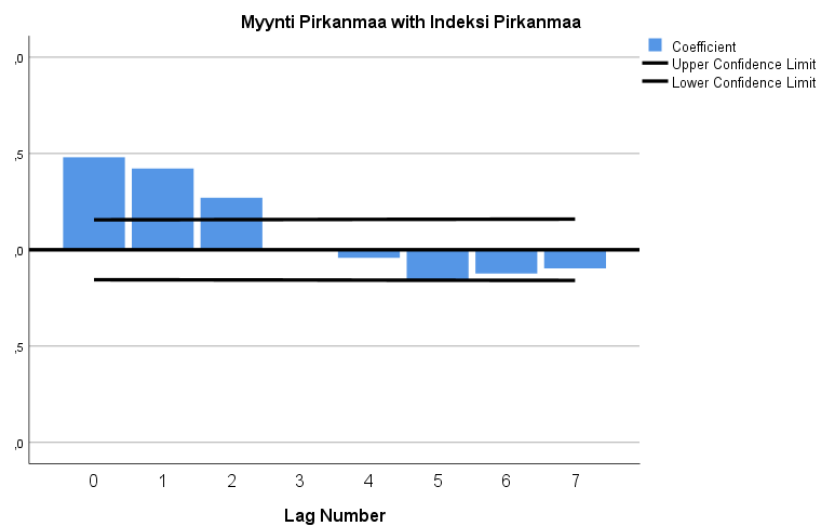


Cross Correlations

Series Pair: Myynti Pirkanmaa with Indeks Pirkanmaa

Lag	Cross Correlation	Std. Error ^a
-7	-,072	,080
-6	-,067	,079
-5	-,046	,079
-4	,062	,079
-3	,174	,079
-2	,400	,078
-1	,440	,078
0	,480	,078
1	,422	,078
2	,270	,078
3	,004	,079
4	-,042	,079
5	-,153	,079
6	-,123	,079
7	-,097	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.



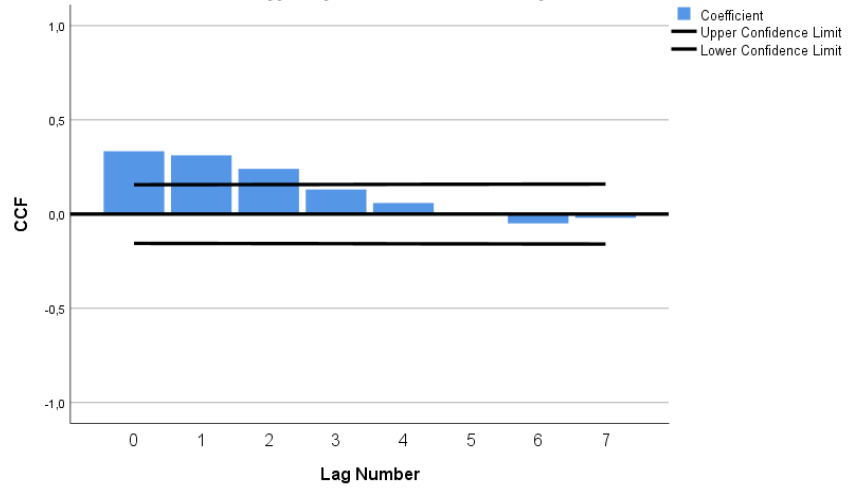
Cross Correlations

Series Pair: Myynti Kymenlaakso with Indeks Kyme

Lag	Cross Correlation	Std. Error ^a
-7	-,059	,080
-6	,023	,079
-5	-,058	,079
-4	,036	,079
-3	,102	,079
-2	,186	,078
-1	,298	,078
0	,334	,078
1	,311	,078
2	,240	,078
3	,131	,079
4	,059	,079
5	,006	,079
6	-,050	,079
7	-,020	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Myynti Kymenlaakso with Indeks Kymenlaakso



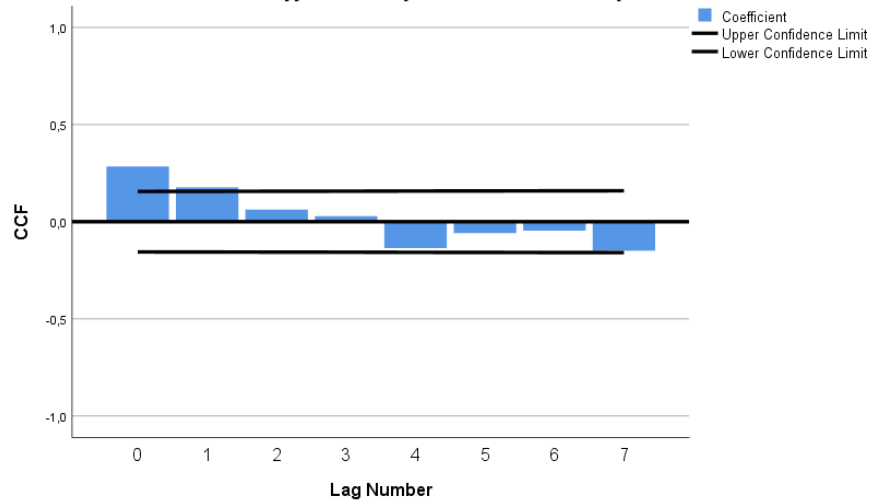
Cross Correlations

Series Pair: Myynti Etelä-Karjala with Indeks Etelä-K

Lag	Cross Correlation	Std. Error ^a
-7	,070	,080
-6	,059	,079
-5	-,003	,079
-4	,087	,079
-3	,080	,079
-2	,284	,078
-1	,320	,078
0	,284	,078
1	,177	,078
2	,062	,078
3	,028	,079
4	-,136	,079
5	-,059	,079
6	-,046	,079
7	-,149	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Myynti Etelä-Karjala with Indeks Etelä-Karjala



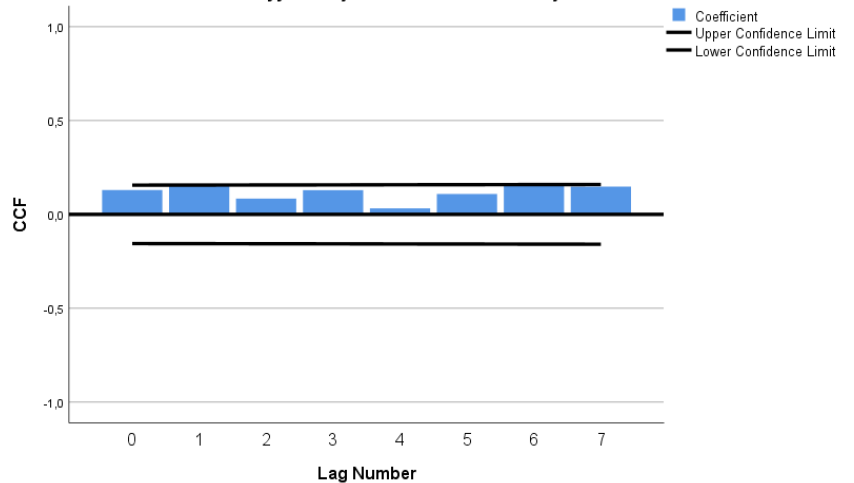
Cross Correlations

Series Pair: Myynti Pohjois-Savo with Indeks Pohjois-S

Lag	Cross Correlation	Std. Error ^a
-7	-,059	,080
-6	-,040	,079
-5	-,037	,079
-4	-,050	,079
-3	,019	,079
-2	,082	,078
-1	,121	,078
0	,129	,078
1	,151	,078
2	,084	,078
3	,129	,079
4	,032	,079
5	,108	,079
6	,163	,079
7	,148	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Myynti Pohjois-Savo with Indeks Pohjois-Savo

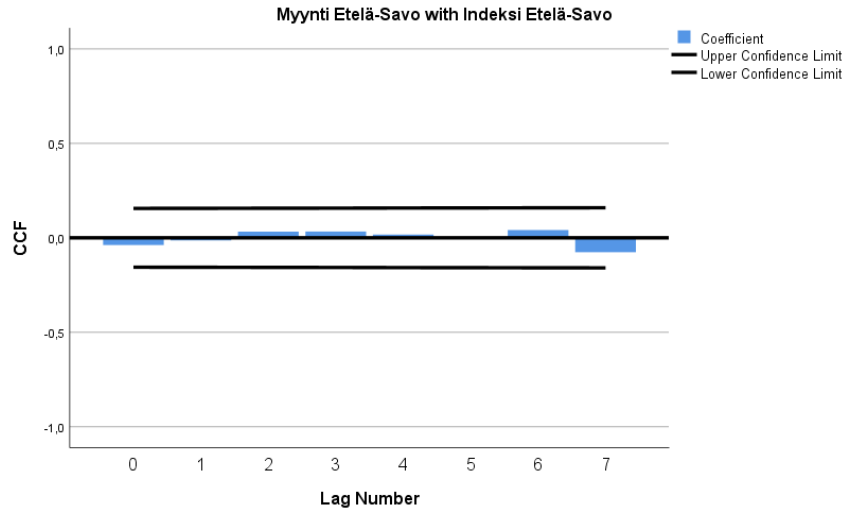


Cross Correlations

Series Pair: Myynti Etelä-Savo with Indeks Etelä-Savo

Lag	Cross Correlation	Std. Error ^a
-7	-,091	,080
-6	-,084	,079
-5	-,074	,079
-4	-,043	,079
-3	-,077	,079
-2	,042	,078
-1	,077	,078
0	-,038	,078
1	-,014	,078
2	,033	,078
3	,034	,079
4	,017	,079
5	-,001	,079
6	,041	,079
7	-,076	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

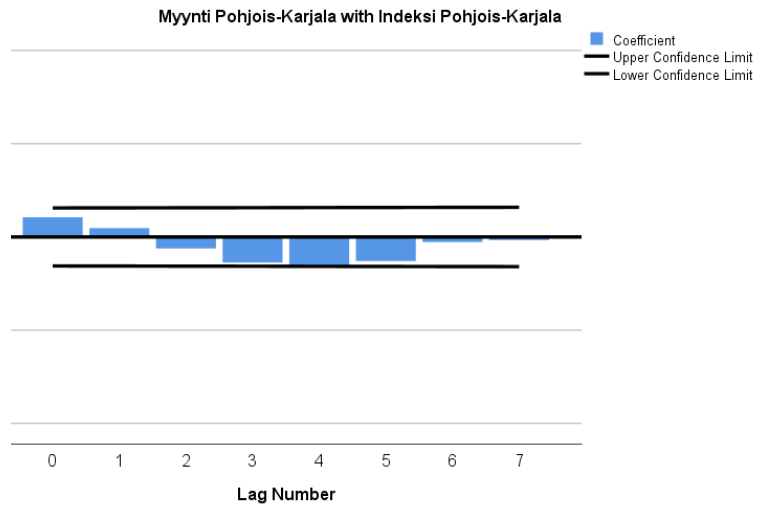


Cross Correlations

Series Pair: Myynti Pohjois-Karjala with Indeks Pohjois-Karjala

Lag	Cross Correlation	Std. Error ^a
-7	,053	,080
-6	,030	,079
-5	,008	,079
-4	,095	,079
-3	,048	,079
-2	,122	,078
-1	,122	,078
0	,106	,078
1	,047	,078
2	-,062	,078
3	-,137	,079
4	-,154	,079
5	-,129	,079
6	-,027	,079
7	-,017	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

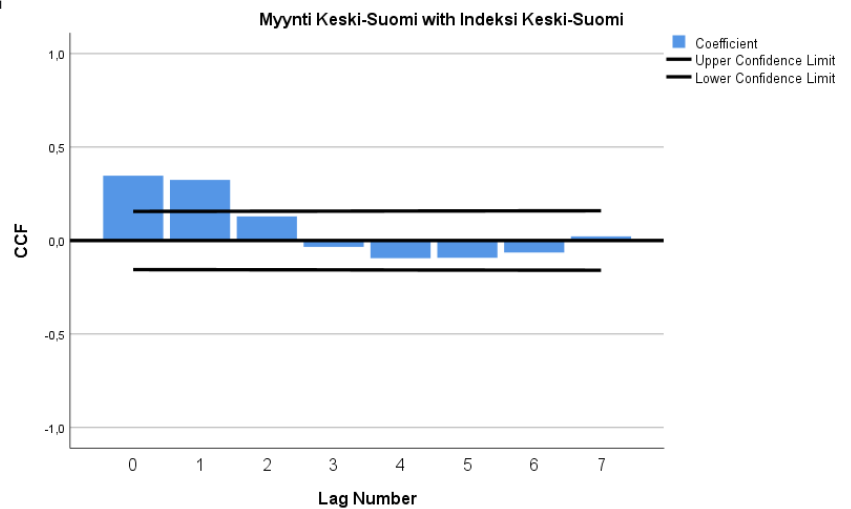


Cross Correlations

Series Pair: Myynti Keski-Suomi with Indeks Keski-Suomi

Lag	Cross Correlation	Std. Error ^a
-7	-,048	,080
-6	-,070	,079
-5	-,165	,079
-4	-,090	,079
-3	,054	,079
-2	,218	,078
-1	,295	,078
0	,347	,078
1	,324	,078
2	,129	,078
3	-,034	,079
4	-,094	,079
5	-,092	,079
6	-,064	,079
7	,022	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

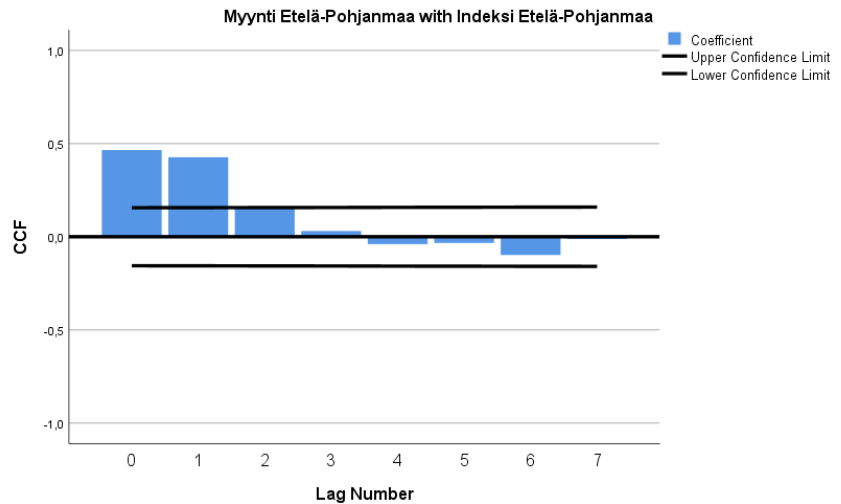


Cross Correlations

Series Pair: Myynti Etelä-Pohjanmaa with Indeks Etelä-Pohjanmaa

Lag	Cross Correlation	Std. Error ^a
-7	-,106	,080
-6	-,129	,079
-5	-,086	,079
-4	-,070	,079
-3	,158	,079
-2	,361	,078
-1	,478	,078
0	,465	,078
1	,427	,078
2	,152	,078
3	,031	,079
4	-,039	,079
5	-,033	,079
6	-,098	,079
7	-,012	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

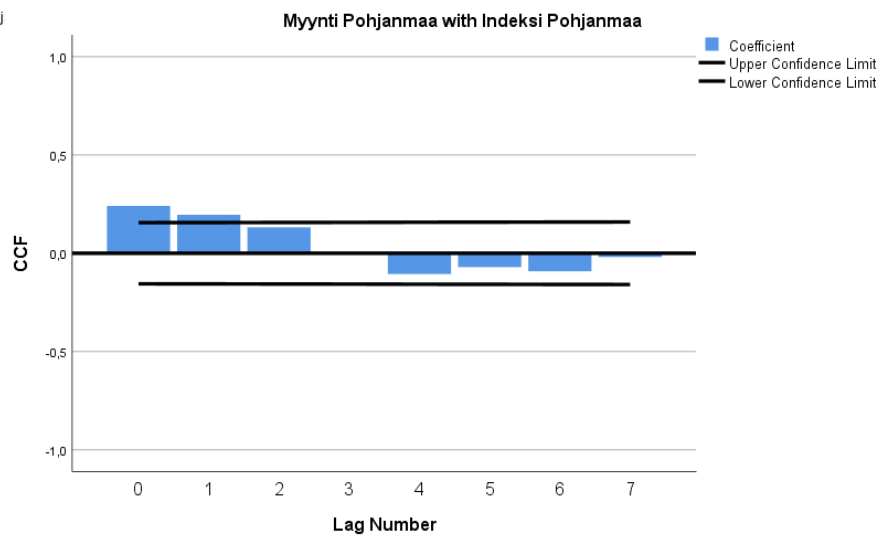


Cross Correlations

Series Pair: Myynti Pohjanmaa with Indeks Pohj

Lag	Cross Correlation	Std. Error ^a
-7	-,126	,080
-6	-,087	,079
-5	-,071	,079
-4	,033	,079
-3	,118	,079
-2	,216	,078
-1	,345	,078
0	,240	,078
1	,196	,078
2	,131	,078
3	-,004	,079
4	-,106	,079
5	-,071	,079
6	-,092	,079
7	-,019	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

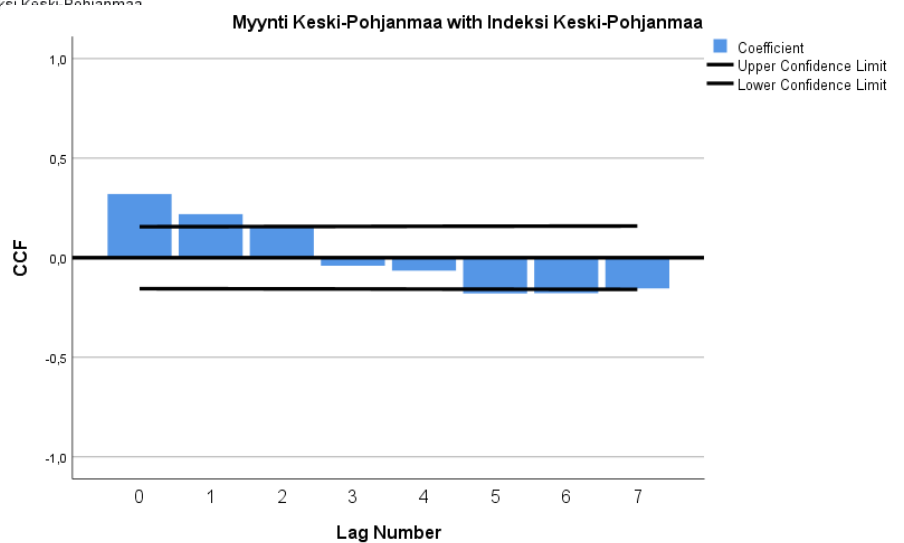


Cross Correlations

Series Pair: Myynti Keski-Pohjanmaa with Indeks Keski-Pohjanmaa

Lag	Cross Correlation	Std. Error ^a
-7	-,123	,080
-6	-,106	,079
-5	-,056	,079
-4	-,028	,079
-3	,106	,079
-2	,202	,078
-1	,240	,078
0	,320	,078
1	,219	,078
2	,157	,078
3	-,040	,079
4	-,065	,079
5	-,180	,079
6	-,178	,079
7	-,154	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.



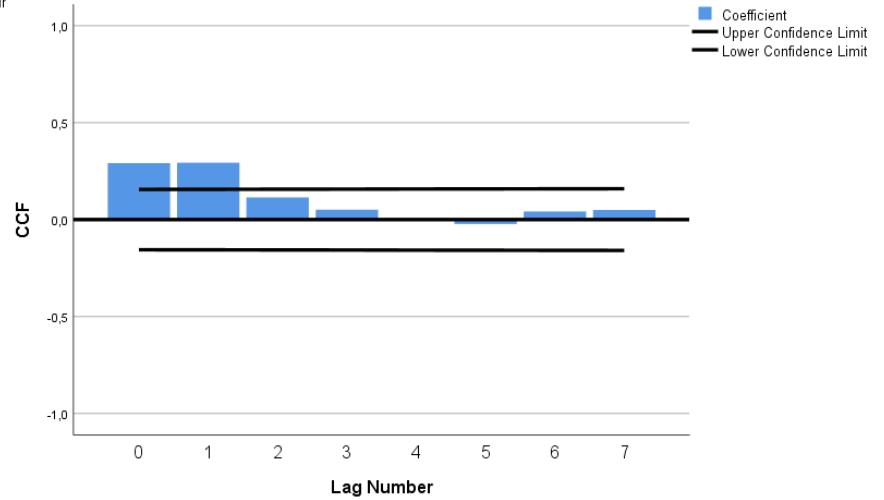
Cross Correlations

Series Pair: Myynti Pohjois-Pohjanmaa with Ir

Lag	Cross Correlation	Std. Error ^a
-7	-,070	,080
-6	-,022	,079
-5	-,033	,079
-4	-,004	,079
-3	,143	,079
-2	,299	,078
-1	,282	,078
0	,291	,078
1	,293	,078
2	,114	,078
3	,051	,079
4	,007	,079
5	-,023	,079
6	,042	,079
7	,050	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Myynti Pohjois-Pohjanmaa with Indeksi Pohjois-Pohjanmaa



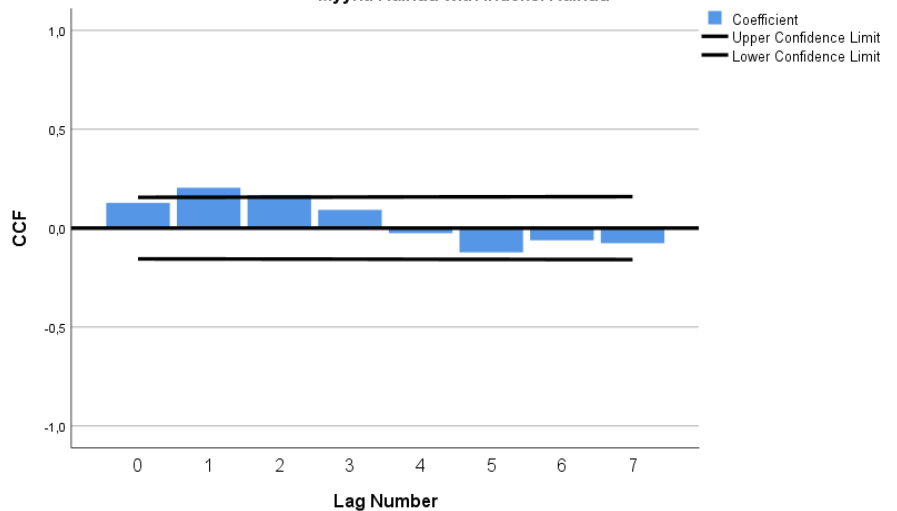
Cross Correlations

Series Pair: Myynti Kainuu with Indeksi Ka

Lag	Cross Correlation	Std. Error ^a
-7	-,151	,080
-6	-,132	,079
-5	-,140	,079
-4	-,131	,079
-3	-,150	,079
-2	-,086	,078
-1	,046	,078
0	,128	,078
1	,204	,078
2	,167	,078
3	,093	,079
4	-,025	,079
5	-,122	,079
6	-,062	,079
7	-,076	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Myynti Kainuu with Indeksi Kainuu



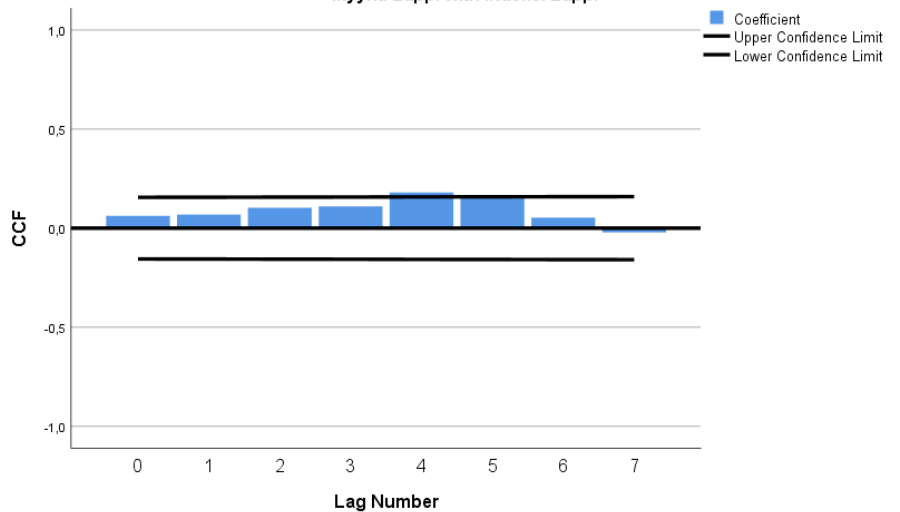
Cross Correlations

Series Pair: Myynti Lappi with Indeksi Lappi

Lag	Cross Correlation	Std. Error ^a
-7	-,036	,080
-6	-,009	,079
-5	-,025	,079
-4	,057	,079
-3	,113	,079
-2	,091	,078
-1	,112	,078
0	,062	,078
1	,069	,078
2	,103	,078
3	,110	,079
4	,180	,079
5	,155	,079
6	,053	,079
7	-,021	,080

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Myynti Lappi with Indeksi Lappi



Liite 9. Maakuntien ristikorraatiokertoimet viiveluvuilla 0, 1 ja 2 ja tilastollinen merkitsevyys

Aluekoodien lähde: Tilastokeskus 2019g. Maakuntaluokitus.

Tilastollisesti merkitsevät arvot on merkitty harmaalla Excel-taulukkoon.

	A	B	C	D	E
1	Aluekoodi	Alue	Viive 0	Viive 1	Viive 2
2	MK01	Uusimaa	0,603	0,569	0,368
3	MK02	Varsinais-Suomi	0,352	0,303	0,152
4	MK04	Satakunta	0,225	0,160	0,092
5	MK05	Kanta-Häme	0,334	0,261	0,155
6	MK06	Pirkanmaa	0,480	0,422	0,270
7	MK07	Päijät-Häme	0,203	0,165	0,119
8	MK08	Kymenlaakso	0,334	0,311	0,240
9	MK09	Etelä-Karjala	0,284	0,177	0,062
10	MK10	Etelä-Savo	-0,038	-0,014	0,033
11	MK11	Pohjois-Savo	0,129	0,151	0,084
12	MK12	Pohjois-Karjala	0,106	0,047	-0,062
13	MK13	Keski-Suomi	0,347	0,324	0,129
14	MK14	Etelä-Pohjanmaa	0,465	0,427	0,152
15	MK15	Pohjanmaa	0,240	0,196	0,131
16	MK16	Keski-Pohjanmaa	0,320	0,219	0,157
17	MK17	Pohjois-Pohjanmaa	0,291	0,293	0,114
18	MK18	Kainuu	0,128	0,204	0,167
19	MK19	Lappi	0,062	0,069	0,103

Liite 10. Maakuntien ristikorrelaatiokertoimien luokittelu karttojen tekemistä varten

Aluekoodien lähde: Tilastokeskus 2019g. Maakuntaluokitus.

Excel-taulukon selitykset:

- 2 = Tilastollisesti merkitsevä positiivinen korrelaatio (harmaa)
- 1 = Tilastollisesti merkitsevästi ei korrelaatiota (harmaa)
- 3 = Ei-tilastollisesti merkitsevä arvo

	A	B	C	D	E
1	Aluekoodi	Alue	Viive 0	Viive 1	Viive 2
2	MK01	Uusimaa	2	2	2
3	MK02	Varsinais-Suomi	2	2	1
4	MK04	Satakunta	1	1	3
5	MK05	Kanta-Häme	2	1	1
6	MK06	Pirkanmaa	2	2	1
7	MK07	Päijät-Häme	1	1	3
8	MK08	Kymenlaakso	2	2	1
9	MK09	Etelä-Karjala	1	1	3
10	MK10	Etelä-Savo	3	3	3
11	MK11	Pohjois-Savo	3	1	3
12	MK12	Pohjois-Karjala	3	3	3
13	MK13	Keski-Suomi	2	2	3
14	MK14	Etelä-Pohjanmaa	2	2	1
15	MK15	Pohjanmaa	1	1	3
16	MK16	Keski-Pohjanmaa	2	1	1
17	MK17	Pohjois-Pohjanmaa	1	1	3
18	MK18	Kainuu	3	1	1
19	MK19	Lappi	3	3	3

Liite 11. Pearsonin korrelaatiomatriisi

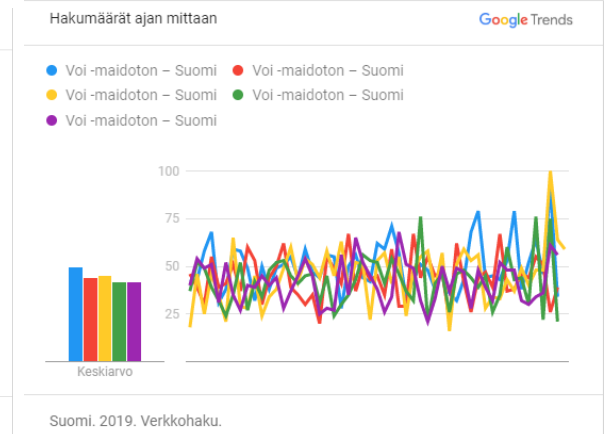
SPSS 2020

Correlations														
		Indeksi Suomi	Indeksi Uusimaa	Indeksi Varsinais- Suomi	Indeksi Satakunta	Indeksi Kanta- Häme	Indeksi Pirkanmaa	Indeksi Päijät- Häme	Indeksi Kymenlaakso	Indeksi Etelä- Karjala	Indeksi Etelä- Savo	Indeksi Pohjois- Savo	Indeksi Pohjois- Karjala	Indeksi Keski- Suomi
Indeksi Suomi	Pearson Correlation	1	,835**	,399**	,312**	,306**	,611**	,392**	,306**	,336**	,147	,191*	,174*	,436**
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,060	,014	,025	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Uusimaa	Pearson Correlation	,835**	1	,382**	,305**	,274**	,621**	,344**	,242**	,304**	,114	,156*	,096	,411**
	Sig. (2-tailed)	,000		,000	,000	,000	,000	,000	,002	,000	,143	,046	,222	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Varsinais-Suomi	Pearson Correlation	,399**	,382**	1	,174*	,006	,244**	,322**	,150	,010	,064	,033	,035	,145
	Sig. (2-tailed)	,000	,000		,025	,943	,002	,000	,054	,901	,413	,673	,656	,063
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Satakunta	Pearson Correlation	,312**	,305**	,174*	1	,239**	,332**	,070	,095	,128	-,010	,055	,084	,047
	Sig. (2-tailed)	,000	,000	,025		,002	,000	,370	,227	,100	,896	,484	,285	,546
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Kanta-Häme	Pearson Correlation	,306**	,274**	,006	,239**	1	,234**	,103	,011	,187*	-,001	,198*	,093	,146
	Sig. (2-tailed)	,000	,000	,943	,002		,002	,188	,892	,016	,986	,011	,234	,061
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Pirkanmaa	Pearson Correlation	,611**	,621**	,244**	,332**	,234**	1	,372**	,129	,252**	,021	,083	,073	,381**
	Sig. (2-tailed)	,000	,000	,002	,000	,002		,000	,098	,001	,787	,288	,353	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Päijät-Häme	Pearson Correlation	,392**	,344**	,322**	,070	,103	,372**	1	,119	,134	,238**	,049	,005	,030
	Sig. (2-tailed)	,000	,000	,000	,370	,188	,000		,127	,086	,002	,535	,950	,705
	N	165	165	165	165	165	165	165	165	165	165	165	165	165

		Indeksi Etelä- Pohjanmaa	Indeksi Pohjanmaa	Indeksi Keski- Pohjanmaa	Indeksi Pohjois- Pohjanmaa	Indeksi Kainuu	Indeksi Lappi	Mynti Suomi	Mynti Uusimaa	Mynti Varsinais- Suomi	Mynti Satakunta	Mynti Kanta- Häme	Mynti Pirkanmaa	Mynti Päijät- Häme
Indeksi Suomi	Pearson Correlation	,542**	,383**	,192*	,446**	,105	,149	,701**	,715**	,672**	,692**	,696**	,709**	,648**
	Sig. (2-tailed)	,000	,000	,013	,000	,179	,055	,000	,000	,000	,000	,000	,000	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Uusimaa	Pearson Correlation	,505**	,329**	,166*	,386**	,068	,197*	,576**	,603**	,542**	,558**	,569**	,581**	,516**
	Sig. (2-tailed)	,000	,000	,033	,000	,383	,011	,000	,000	,000	,000	,000	,000	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Varsinais-Suomi	Pearson Correlation	,191*	,089	,098	,247**	,034	-,017	,373**	,384**	,352**	,360**	,367**	,372**	,339**
	Sig. (2-tailed)	,014	,253	,209	,001	,665	,825	,000	,000	,000	,000	,000	,000	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Satakunta	Pearson Correlation	,196*	,142	,005	,158*	,080	,040	,229**	,249**	,211**	,225**	,224**	,235**	,200*
	Sig. (2-tailed)	,011	,069	,945	,043	,307	,611	,003	,001	,006	,004	,004	,002	,010
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Kanta-Häme	Pearson Correlation	,136	-,015	,054	,087	,034	-,033	,335**	,337**	,327**	,332**	,334**	,339**	,319**
	Sig. (2-tailed)	,082	,851	,489	,264	,662	,676	,000	,000	,000	,000	,000	,000	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Pirkanmaa	Pearson Correlation	,291**	,181*	,276**	,262**	,101	,192*	,473**	,484**	,453**	,467**	,475**	,480**	,434**
	Sig. (2-tailed)	,000	,020	,000	,001	,198	,014	,000	,000	,000	,000	,000	,000	,000
	N	165	165	165	165	165	165	165	165	165	165	165	165	165
Indeksi Päijät-Häme	Pearson Correlation	,069	,229**	,140	,189*	,041	,044	,247**	,272**	,219**	,228**	,232**	,245**	,203**
	Sig. (2-tailed)	,381	,003	,074	,015	,605	,577	,001	,000	,005	,003	,003	,001	,009
	N	165	165	165	165	165	165	165	165	165	165	165	165	165

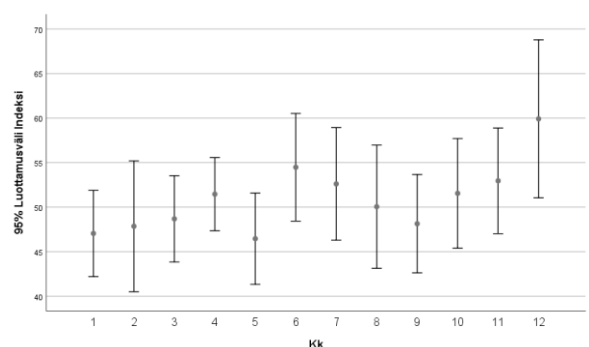
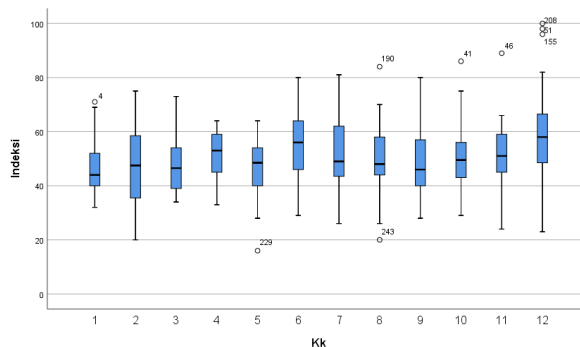
Liite 12. Voin hakuajankohdat

Voin hakemisen trendi 5 vuoden (2015–2019) ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

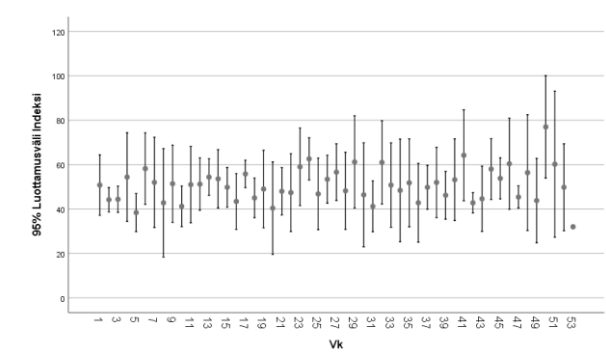
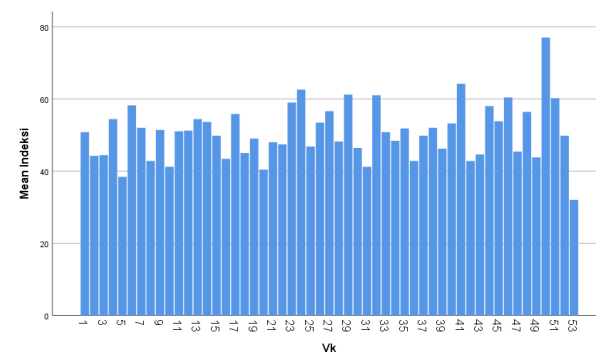


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

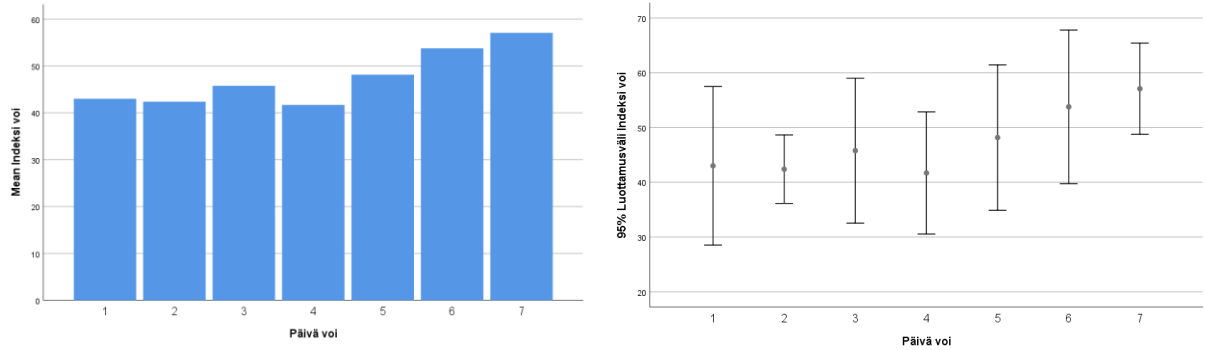
Voin viiden vuoden (2015–2019) hakuindeksien keskiarvojen määrän jakauma kuukausittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Voin Google Trends -hakuindeksien 5 vuoden keskiarvot viikkonumeroittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Voin hakeminen viikonpäivittäin edellisen 90 päivän indeksien keskiarvona (Google Trends 18.2.2020) (vasen) ja luottamusvälit (oikea).

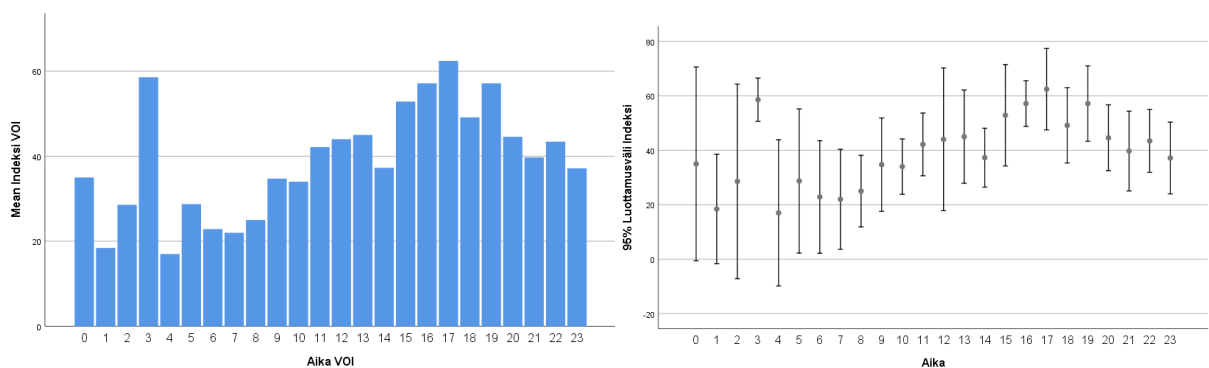


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

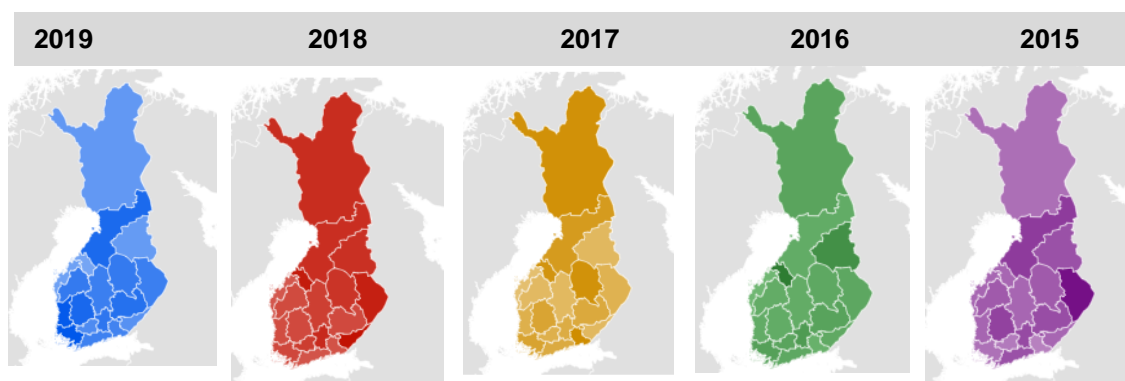
Voin hakeminen seitsemän päivän aikana (Google Trends 22.1.2010)



Kuvio x. Voin hakemisen indeksien kellonaikojen keskiarvot seitsemän päivän aikana (Google Trends 22.1.2020) (vasen) ja luottamusvälit (oikea)

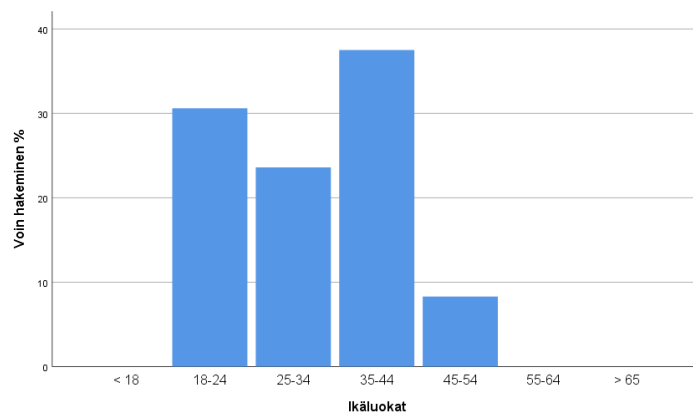


Voin hakemisen alueelliset erot eri vuosina (Google Trends 14.2.2020)



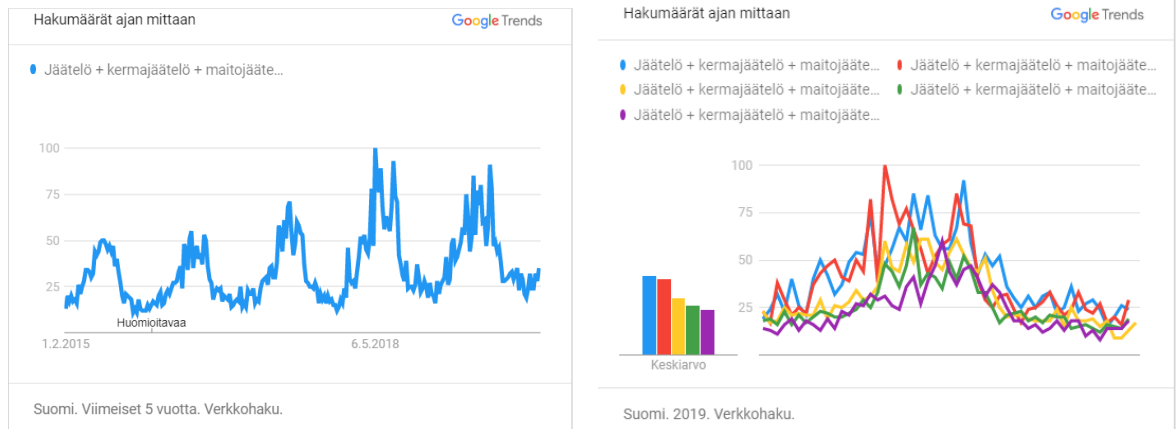
Satakunta	
Keski-Suomi	
Varsinais-Suomi	
Pohjois-Karjala	
Pohjois-Pohjanmaa	
Pirkanmaa	
Päijät-Häme	
Etelä-Pohjanmaa	
Pohjois-Savo	
Etelä-Karjala	
Kanta-Häme	
Uusimaa	
Etelä-Savo	
Pohjanmaa	
Kymenlaakso	
Lappi	
Kainuu	
Keski-Pohjanmaa	

Voin hakeminen ikäluokittain (Ubersuggest 25.1.2020)



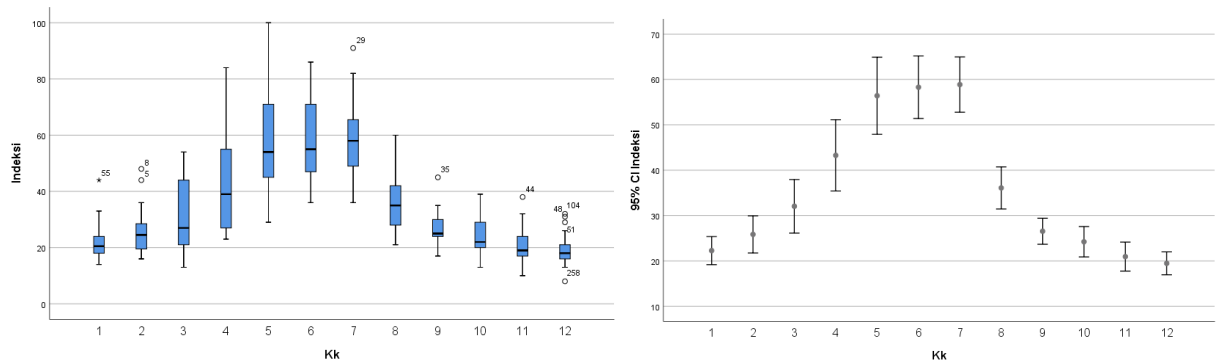
Liite 13. Jäätelön hakuajankohdat

Jäätelön hakemisen trendi 5 vuoden ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosittainen vaihtelu Suomessa vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

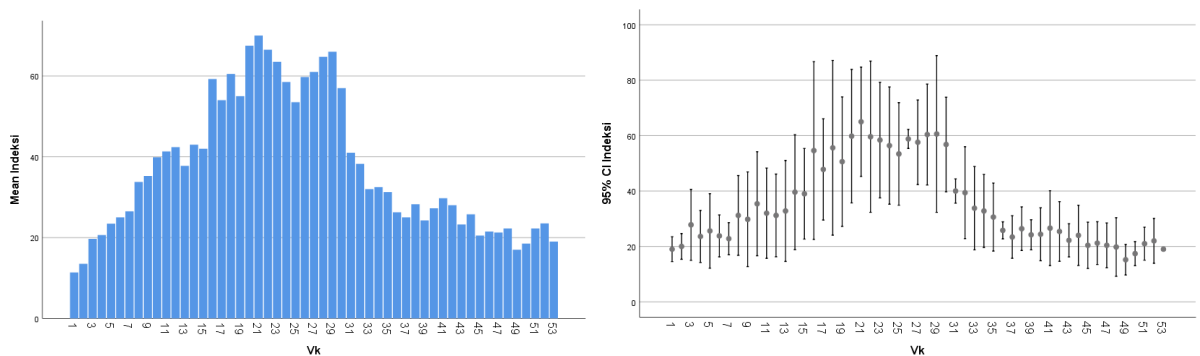


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

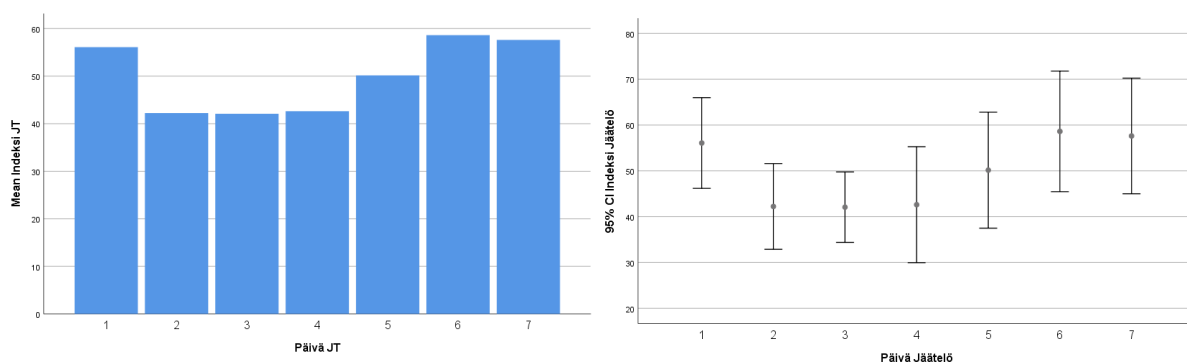
Jäätelön Google Trends -hakuindeksien viiden vuoden (2015–2019) keskiarvojen indeksien määrän jakauma kuukausittain (vasen) ja luottamusvälit (oikea)



Jäätelön Google Trends -hakuindeksien viiden vuoden keskiarvot viikoittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Jäätelön hakeminen viikonpäivittäin 90 päivän keskiarvona (Google Trends 18.2.2020)
(vasen) ja luottamusvälit (oikea)

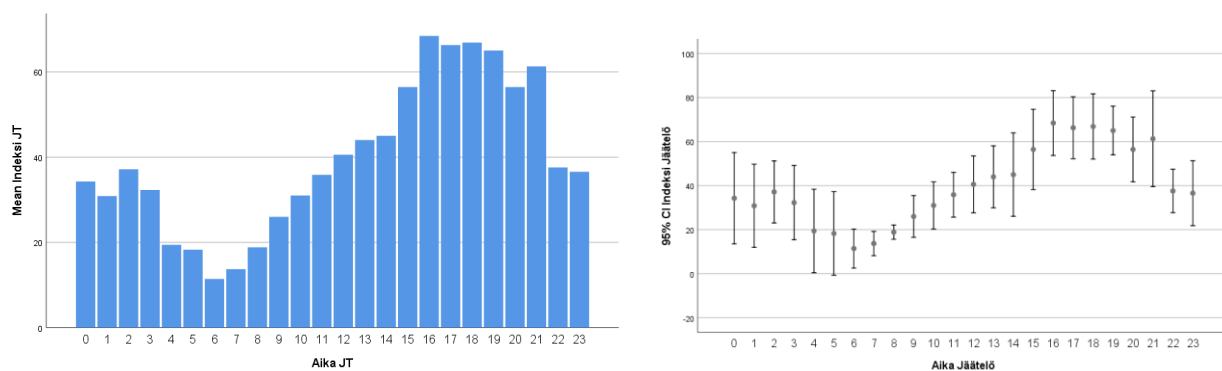


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

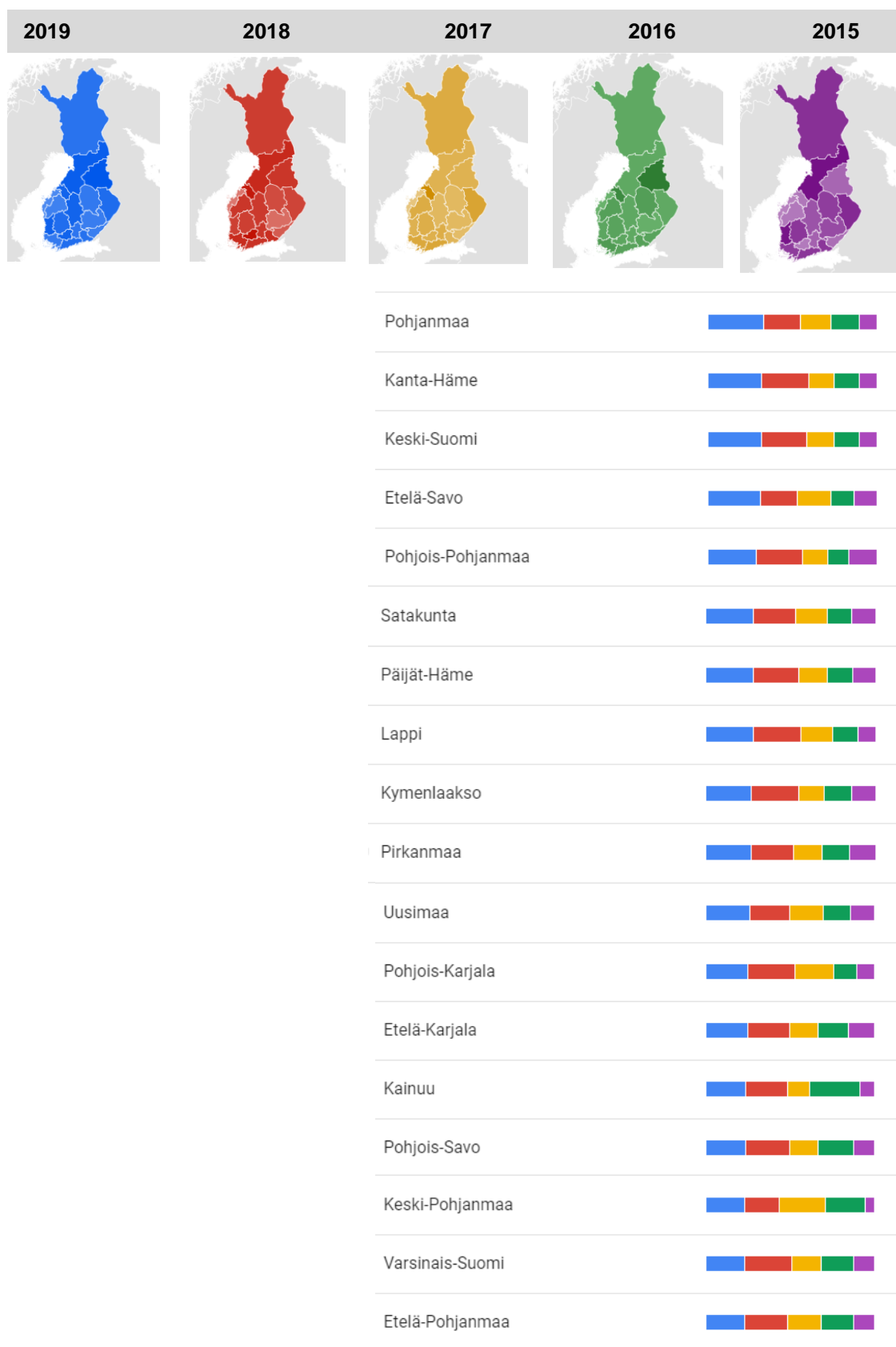
Jäätelön hakeminen seitsemän päivän aikana (Google Trends 22.1.2020)



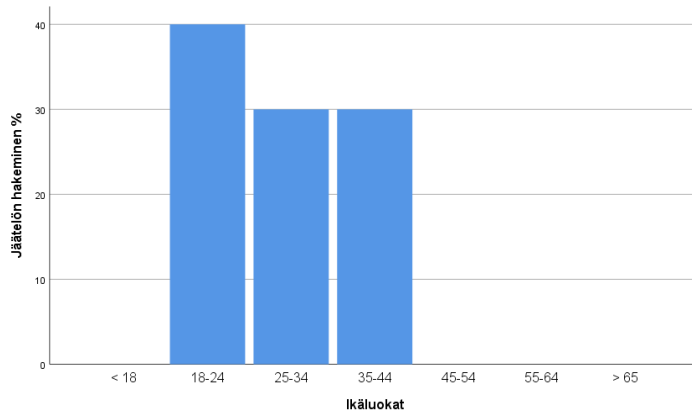
Jäätelön hakemisen indeksien keskiarvot tunneittain viikon aikana (Google Trends 22.1.2020) (vasen) ja luottamusvälit (oikea)



Jäätelön hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)

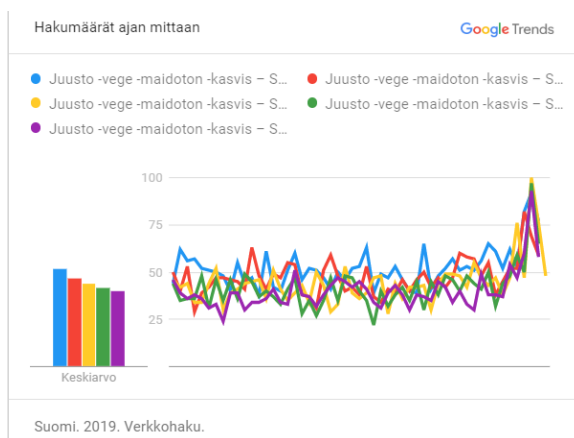


Jäätelön hakeminen ikäluokittain (Ubersuggest 25.1.2020)



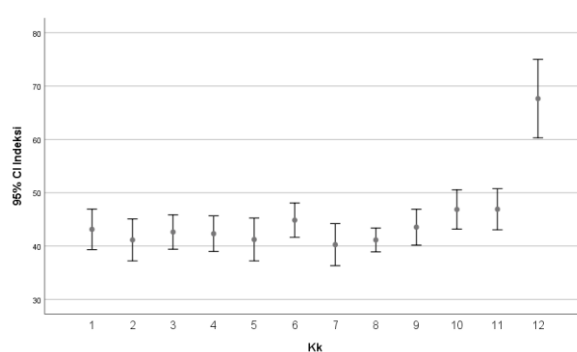
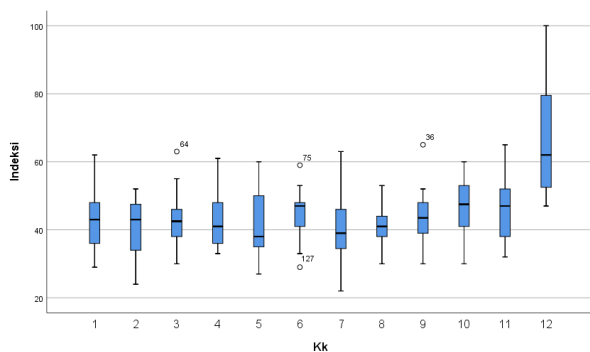
Liite 14. Juuston hakuajankohdat

Juuston hakemisen trendi viiden vuoden ajanjaksolla Suomessa (Google Trends 1.2.2020) (vasen) ja vuosittainen vaihtelu vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

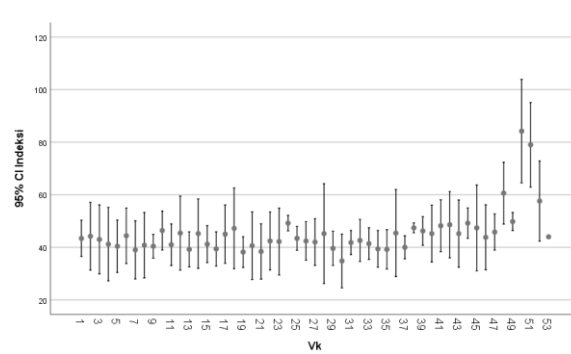
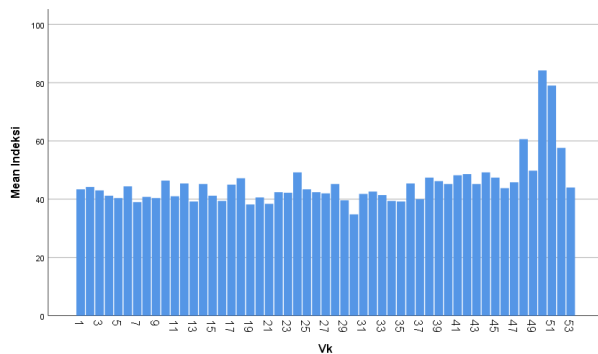


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

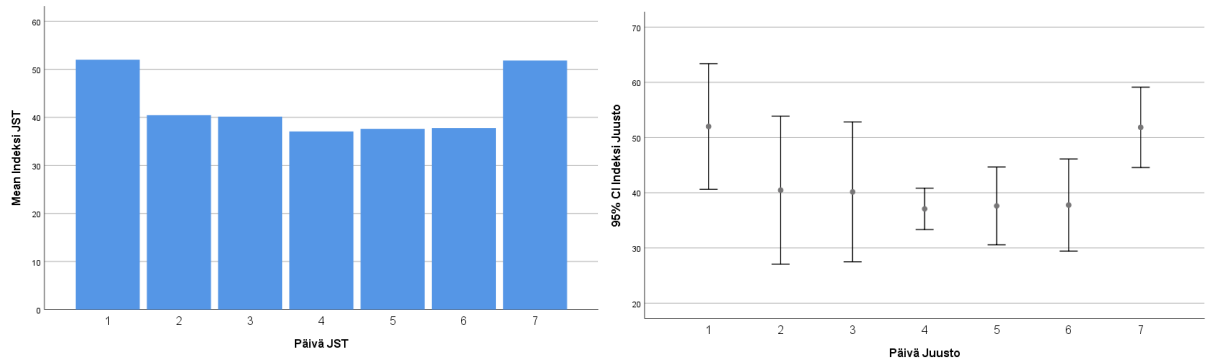
Juuston Google Trends -hakuindeksien viiden vuoden (2015–2019) keskiarvojen indeksien määrän jakauma kuukausittain (vasen) ja luottamusvälit (oikea)



Juuston Google Trends -hakuindeksien 5 vuoden keskiarvot viikoittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Juuston hakeminen viikonpäivittäin 90 päivän keskiarvona (Google Trends 18.2.2020)
(vasen) ja luottamusvälit (oikea)

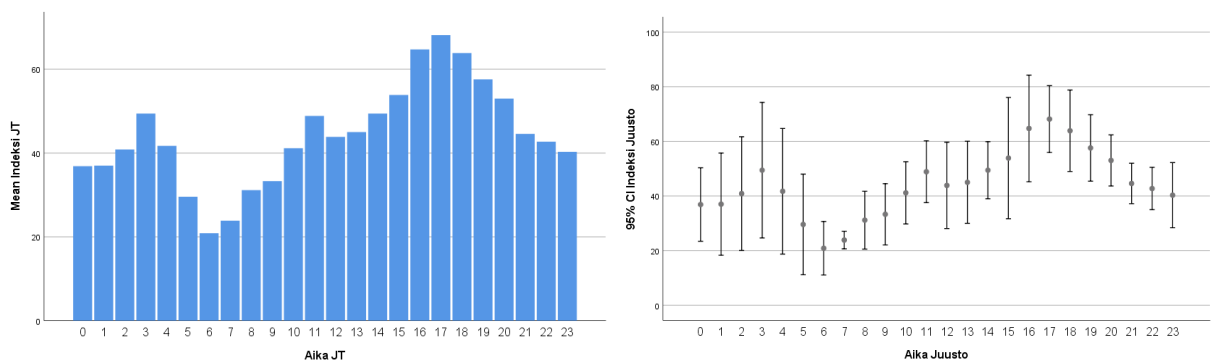


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

Juuston hakeminen viimeisen 7 päivän aikana (Google Trends 22.1.2010)



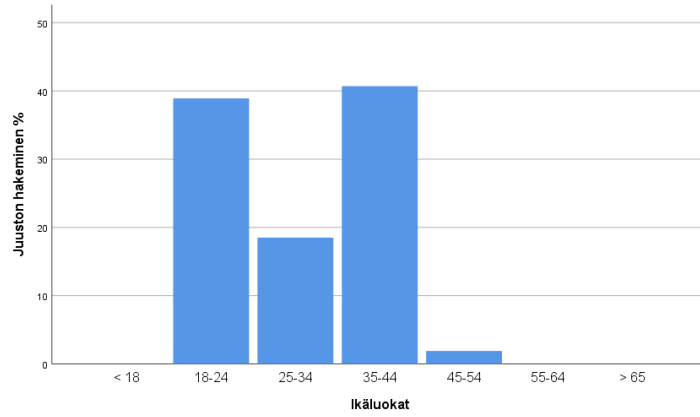
Juuston hakemisen indeksien keskiarvot kellonajoittain viimeisen seitsemän päivän aikana (Google Trends 22.1.2020) (vasen) ja luottamusvälit (oikea)



Juuston hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)

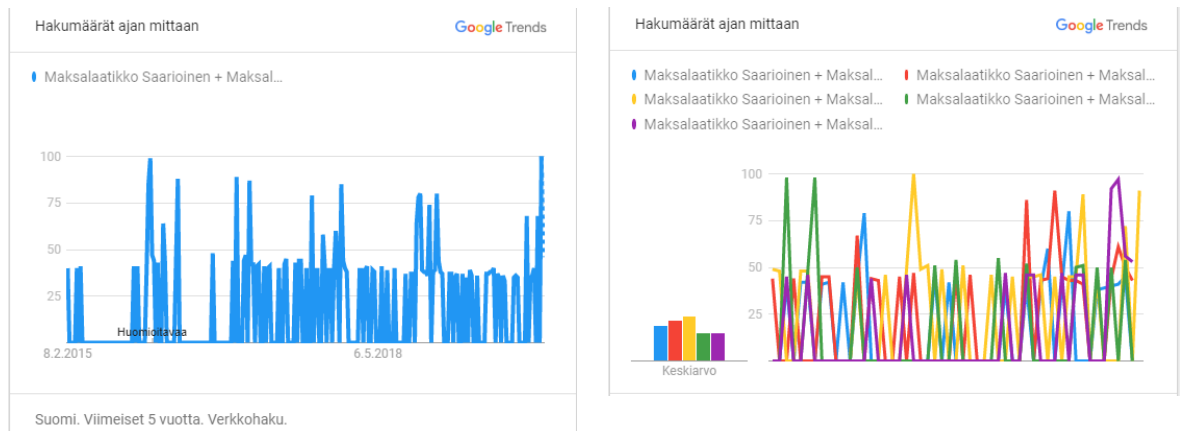


Juuston hakeminen ikäluokittain (Ubersuggest 25.1.2020)



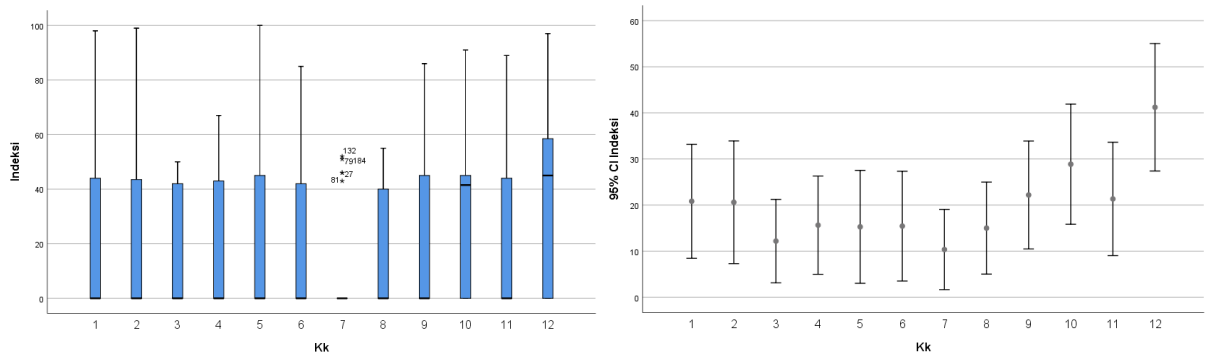
Liite 15. Maksalaatikon hakuajankohdat

Maksalaatikkovalmisruoan hakemisen trendi viiden vuoden ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 14.2.2020) (oikea)

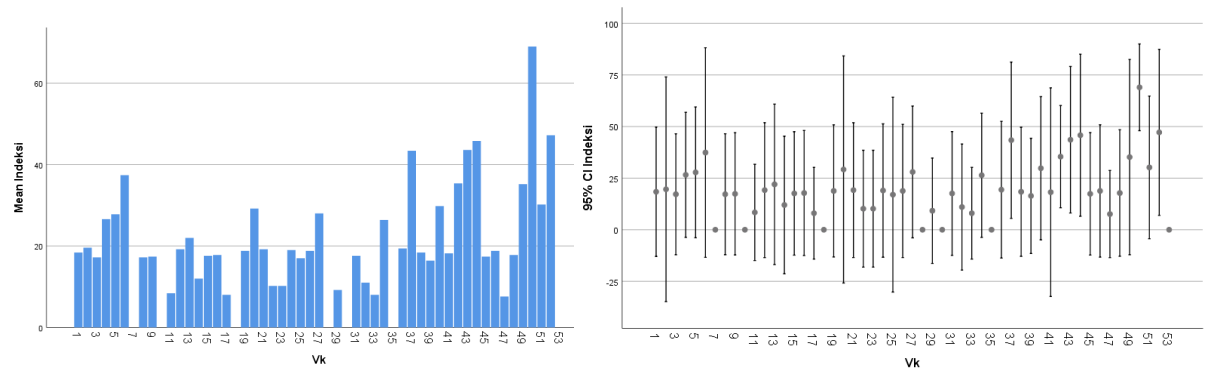


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

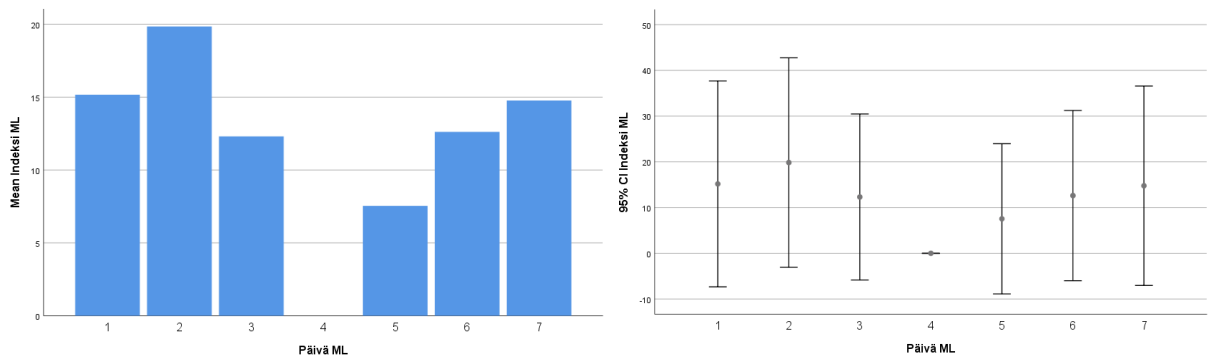
Maksalaatikkovalmisruoan Google Trends -hakuindeksien 5 vuoden (2015–2019) keskiarvojen indeksien määrän jakauma kuukausittain (vasen) ja luottamusvälit (oikea)



Maksalaatikko-valmisruoan Google Trends -hakuindeksien 5 vuoden keskiarvot viikoittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Maksalaatikkovalmisruoan hakeminen viikonpäivittäin 90 päivän keskiarvona (Google Trends 18.2.2020) (vasen) ja luottamusvälit (oikea)

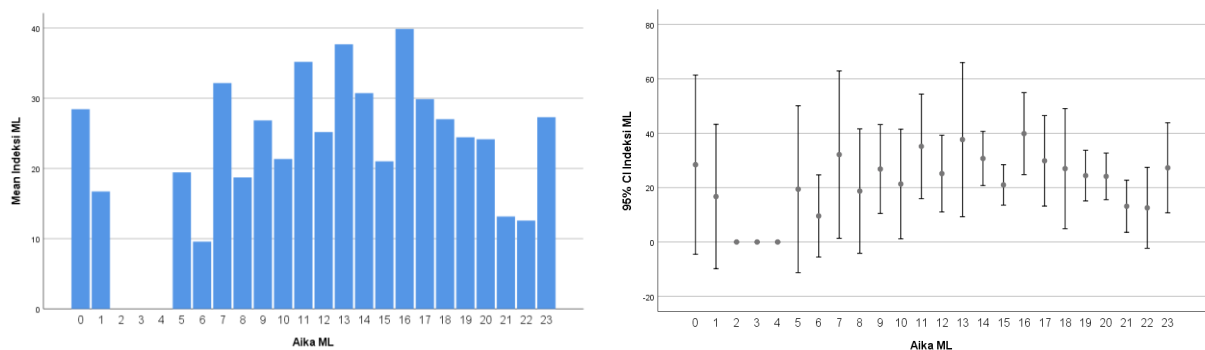


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

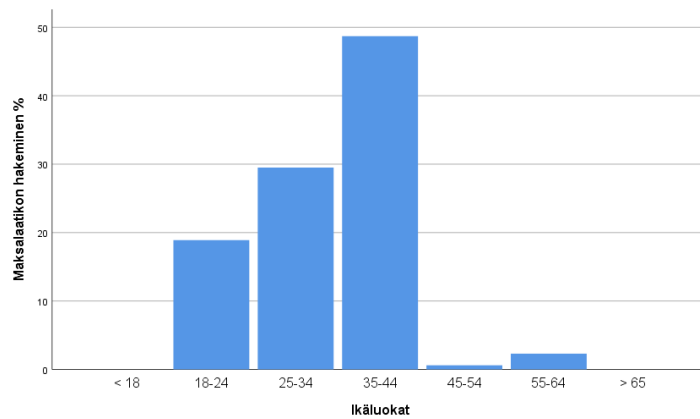
Maksalaatikkovalmisruoan hakeminen viimeisen 7 päivän aikana (Google Trends 22.1.2020)



Maksalaatikko-valmisruoan hakemisen indeksien kellonaikojen keskiarvot seitsemän päivän aikana (Google Trends 22.1.2020) (vasen) ja luottamusvälit (oikea)

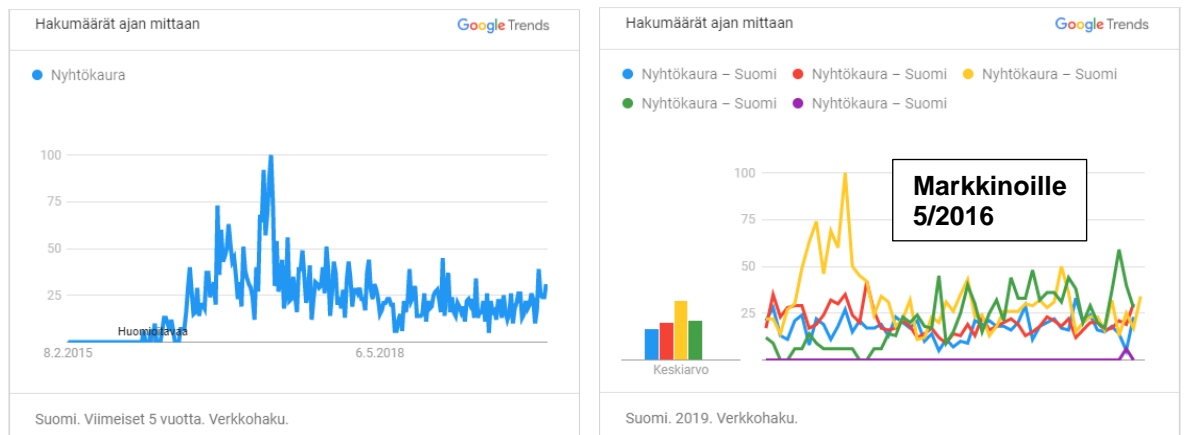


Maksalaatikon hakeminen ikäryhmittäin (Ubersuggest 25.1.2020)



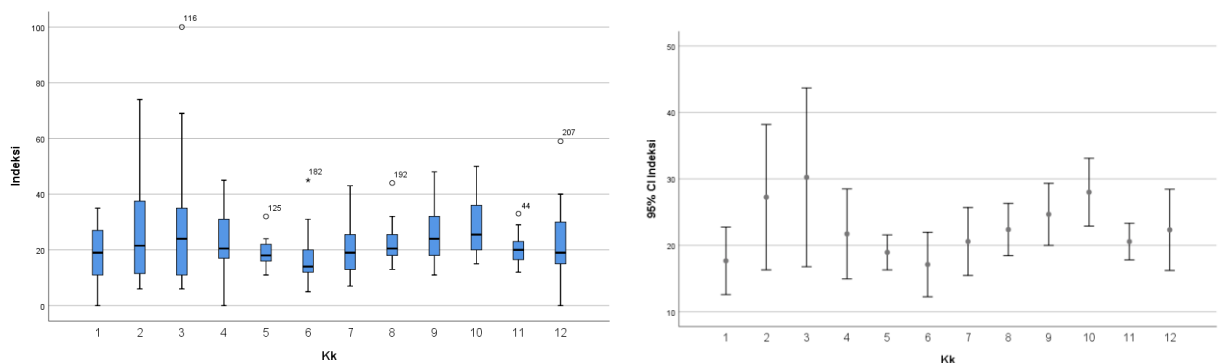
Liite 16. Nyhtökauran hakuajankohdat

Nyhtökauran hakemisen trendi Suomessa viiden vuoden ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

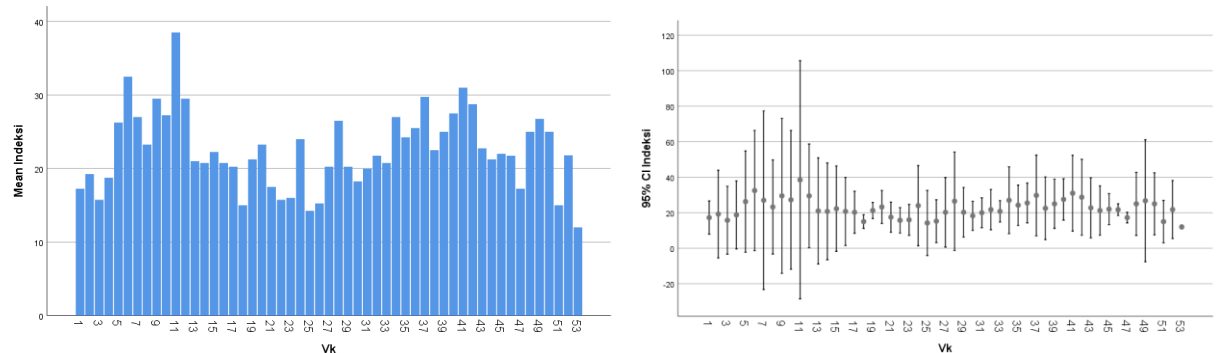


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

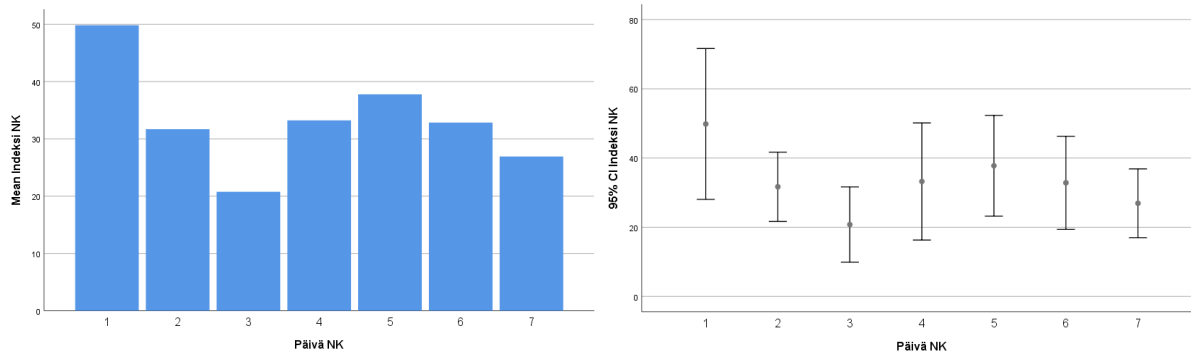
Nyhtökauran Google Trends -hakuindeksien keskiarvojen määrän jakauma kuukausittain ajanjaksolla 5/2016–12/2019 (vasen) ja luottamusvälit (oikea)



Nyhtökauran Google Trends -hakuindeksien keskiarvot viikoittain ajanjaksolla 20.12.2015–31.12.2019 (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Nyhtökauran hakeminen viikonpäivittäin edellisen 90 päivän keskiarvona (Google Trends 18.2.2020) (vasen) ja luottamusvälit (oikea)

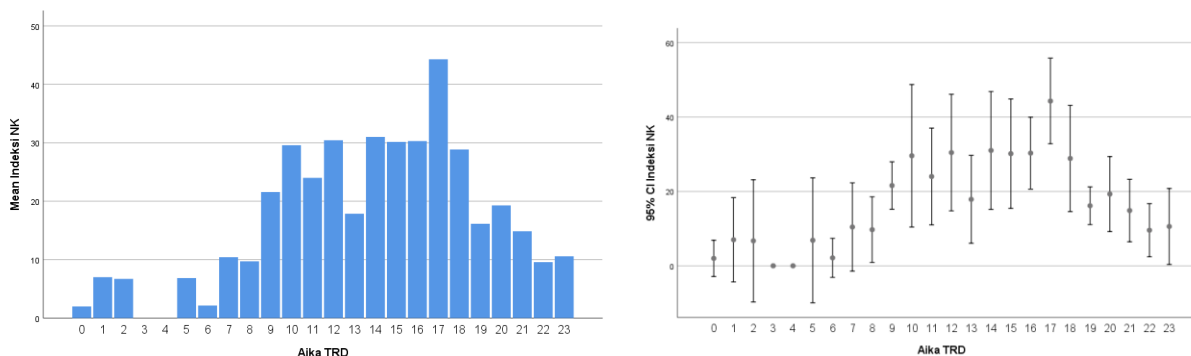


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

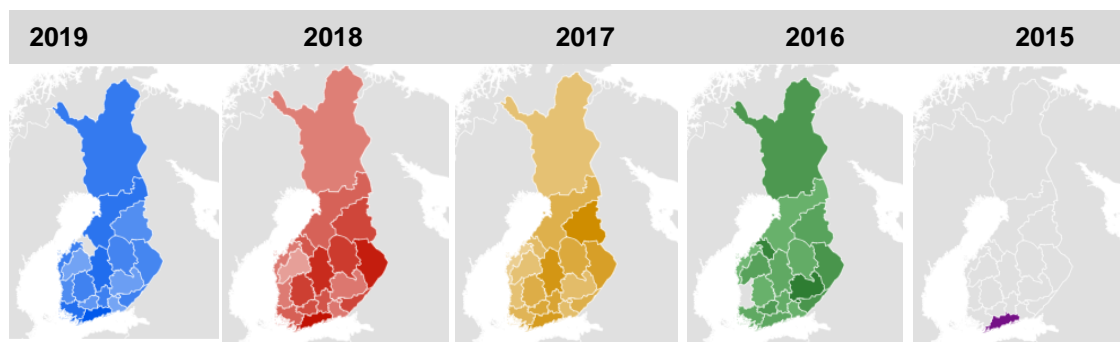
Nyhtökauran hakeminen viimeisen 7 päivän aikana (Google Trends 22.1.2010)



Nyhtökauran hakemisen indeksien seitsemän päivän keskiarvot (Google Trends 22.1.2020) (vasen) ja luottamusvälit (oikea)

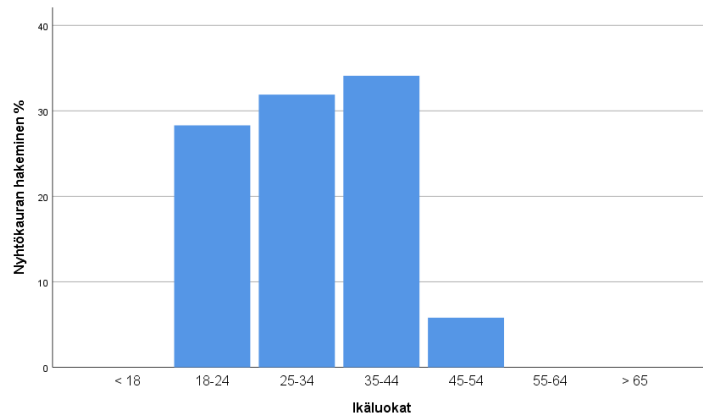


Nyhtökauran aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)



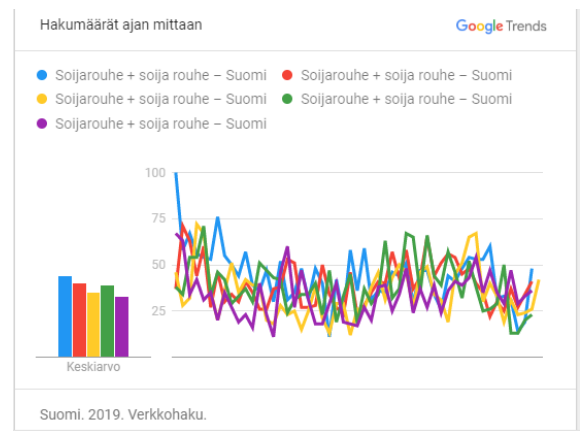
Pohjanmaa	
Pohjois-Pohjanmaa	
Lappi	
Varsinais-Suomi	
Uusimaa	
Päijät-Häme	
Satakunta	
Pirkanmaa	
Keski-Suomi	
Pohjois-Savo	
Etelä-Pohjanmaa	
Kymenlaakso	
Kanta-Häme	
Etelä-Karjala	
Pohjois-Karjala	
Kainuu	
Etelä-Savo	
Keski-Pohjanmaa	

Nyhtökauran hakeminen ikäluokittain (Ubersuggest 25.1.2020)



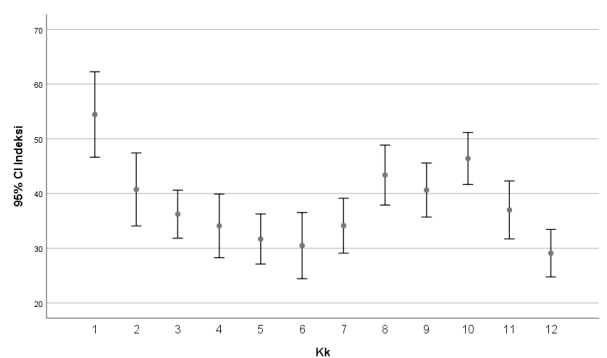
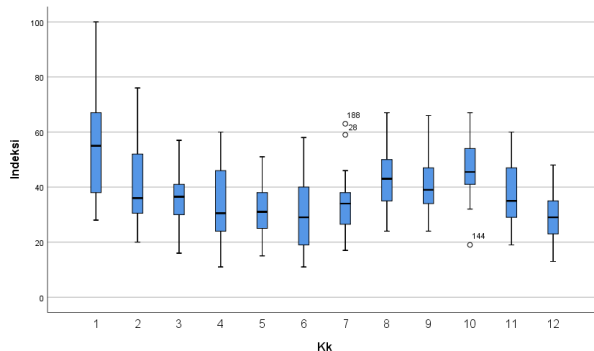
Liite 17. Soijarouheen hakuajankohdat

Soijarouheen hakemisen trendi Suomessa viiden vuoden ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

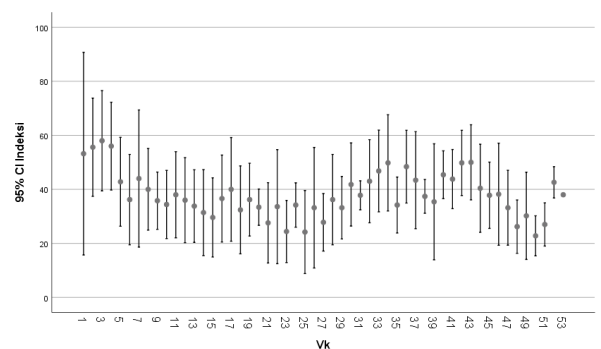
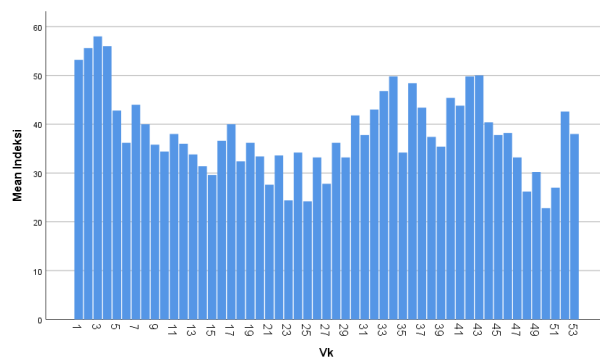


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

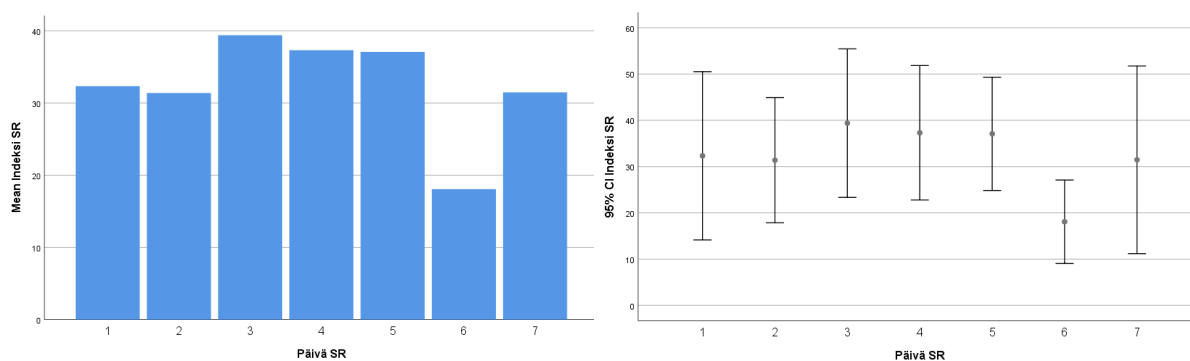
Soijarouheen Google Trends -hakuindeksien 5 vuoden (vuosien 2015–2019) keskiarvojen indeksien määrän jakauma kuukausittain (vasen) ja luottamusvälit (oikea)



Soijarouheen Google Trends -hakuindeksien viiden vuoden (2019–2015) keskiarvot viikoittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Soijarouheen hakeminen viikonpäivittäin edellisen 90 päivän keskiarvona (Google Trends 18.2.2020) (vasen) ja luottamusvälit (oikea)

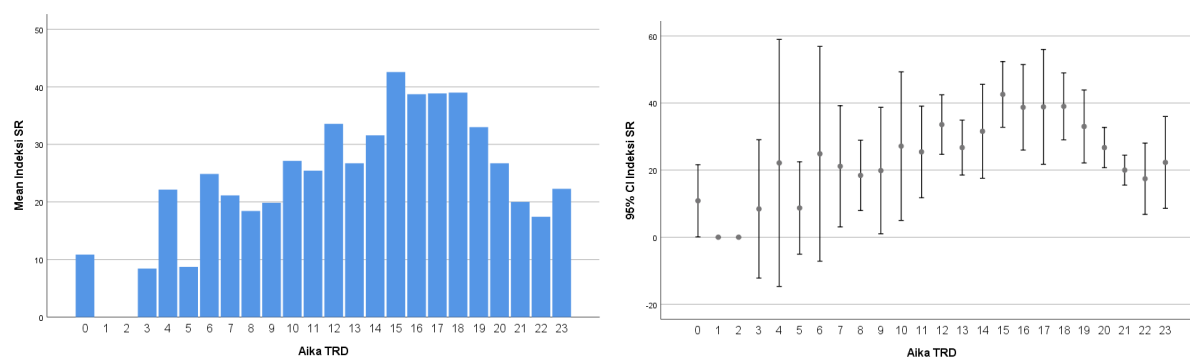


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

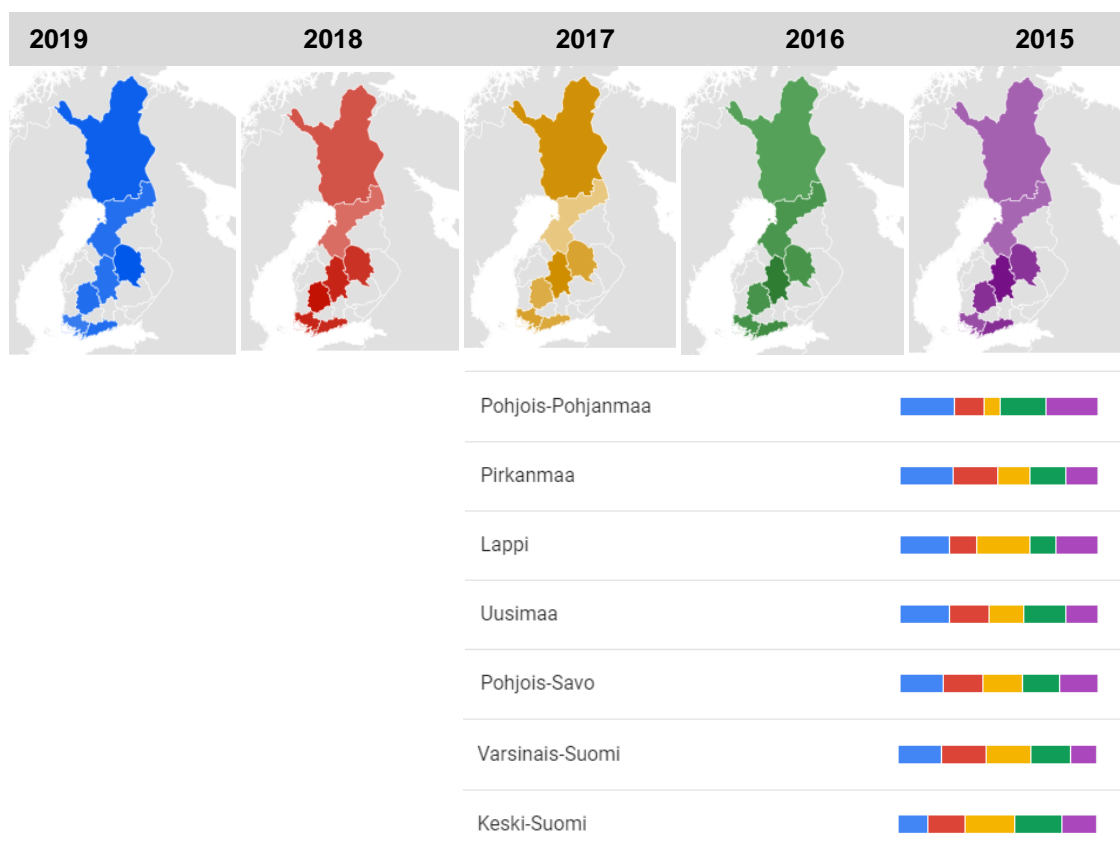
Soijarouheen hakeminen seitsemän päivän aikana (Google Trends 22.1.2010)



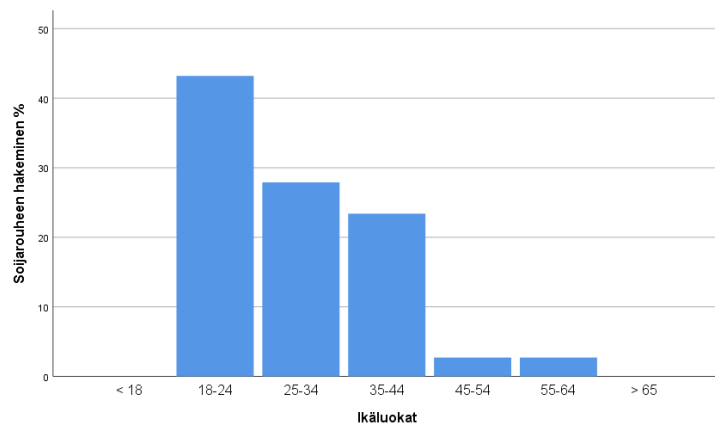
Soijarouheen hakuindeksien kellonaikojen viikon keskiarvot (Google Trends 22.1.2020) (vasen) ja luottamusvälit (oikea)



Soijarouheen hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)

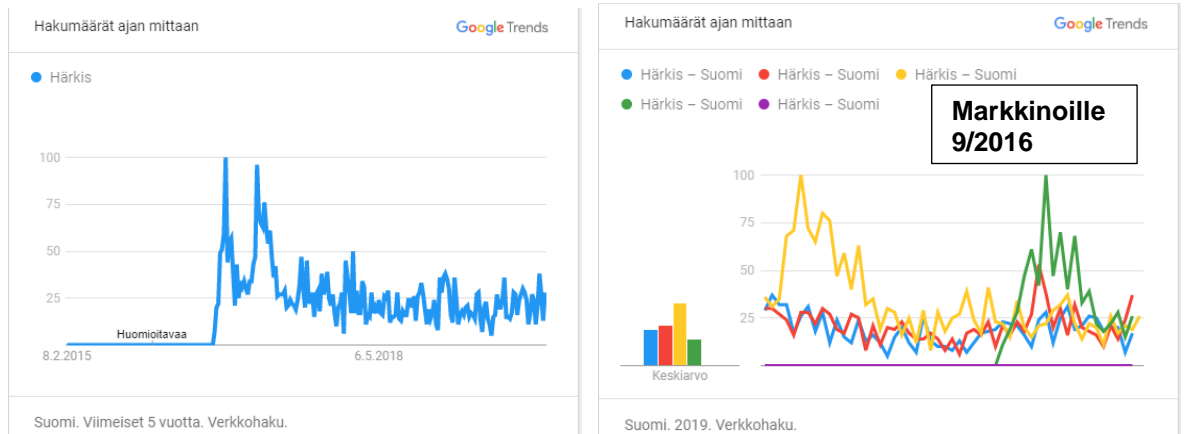


Soijarouheen hakeminen ikäluokittain (Ubersuggest 25.1.2020)



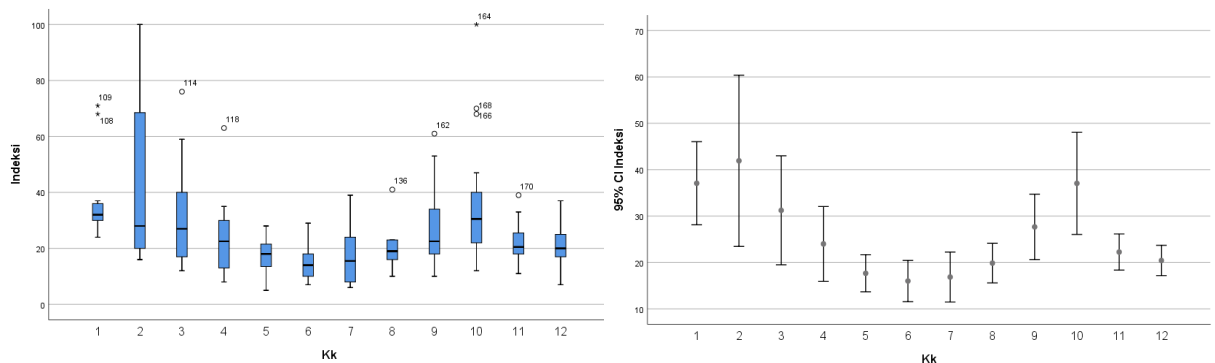
Liite 18. Härkiksen hakuajankohdat

Härkiksen hakemisen trendi Suomessa viiden vuoden ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

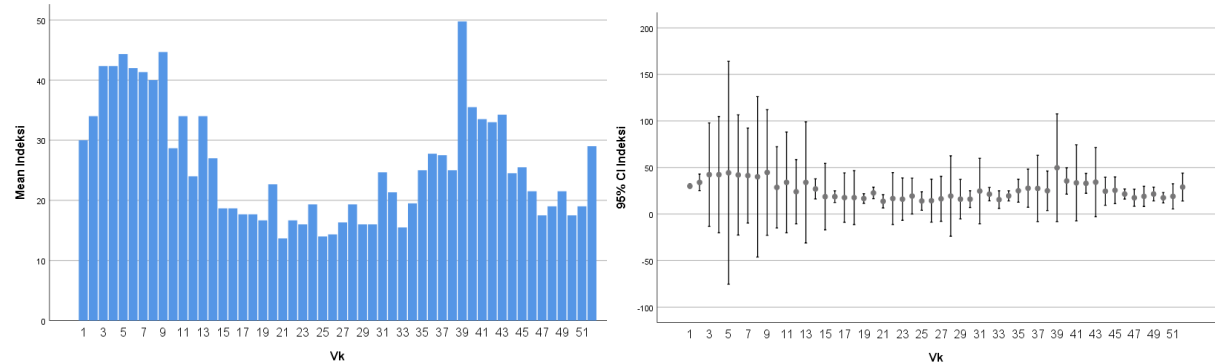


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

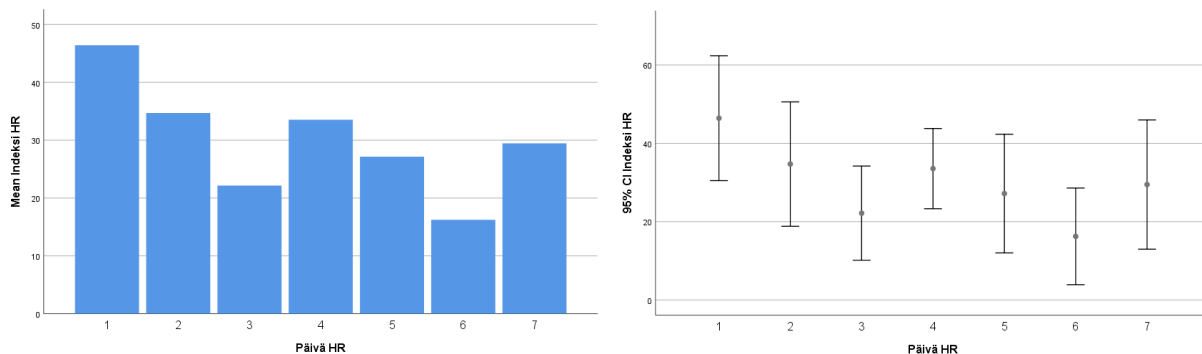
Härkiksen Google Trends -hakuindeksien keskiarvojen määrän jakauma kuukausittain ajanjaksolla 20.8.2016–31.12.2019 (vasen) ja luottamusvälit (oikea)



Härkiksen Google Trends -hakuindeksien keskiarvojen määrän jakauma viikoittain ajanjaksolla 20.8.2016–31.12.2019 (vasen) ja luottamusvälit (oikea)



Härkiksen hakeminen viikonpäivittäin 90 päivän keskiarvona (Google Trends 18.2.2020)
(vasen) ja luottamusvälit (oikea)

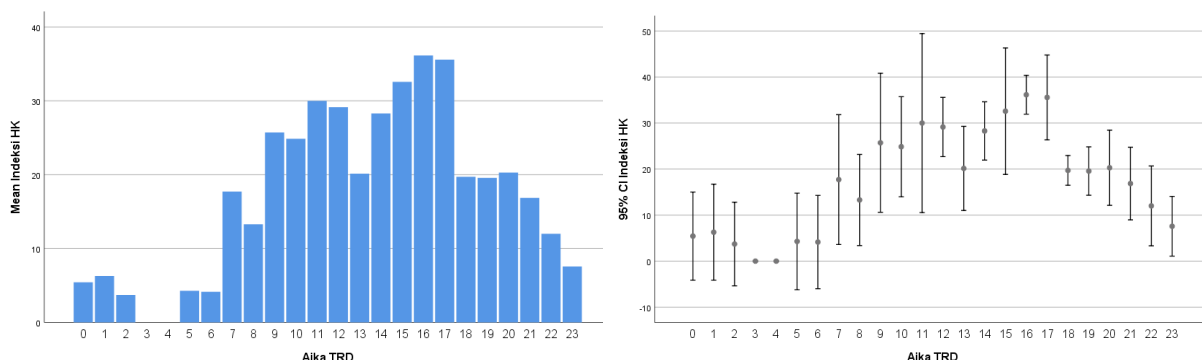


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

Härkiksen hakeminen viimeisen 7 päivän aikana (Google Trends 22.1.2020)



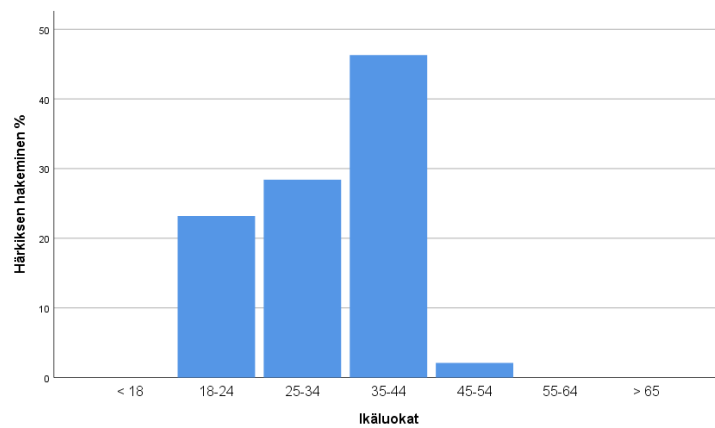
Härkiksen hakuindeksien kellonaikojen keskiarvot viikon ajalta (Google Trends 22.1.2020)
(vasen) ja luottamusvälit (oikea)



Härkiksen hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)

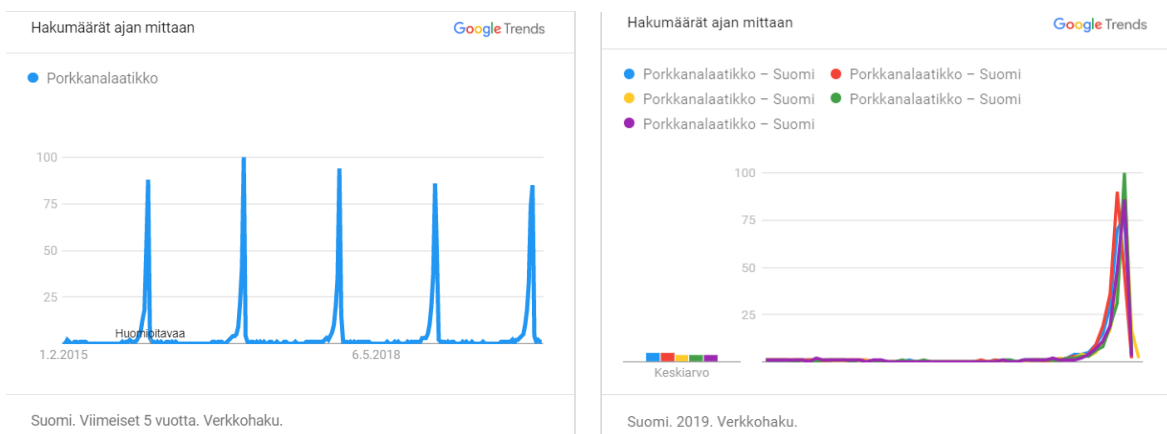


Härkiksen hakeminen ikäluokittain (Ubersuggest 25.1.2020)



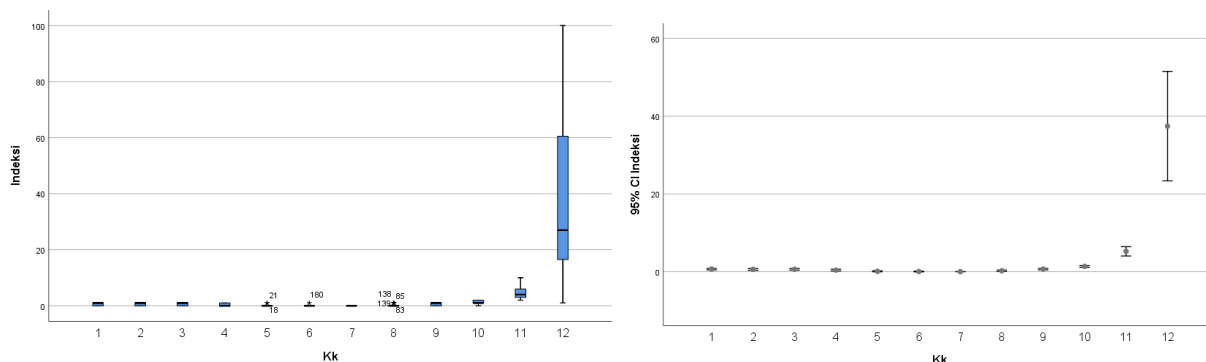
Liite 19. Porkkanalaatikon hakuajankohdat

Porkkanalaatikon hakemisen trendi Suomessa viiden vuoden ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 1.2.2020) (oikea)

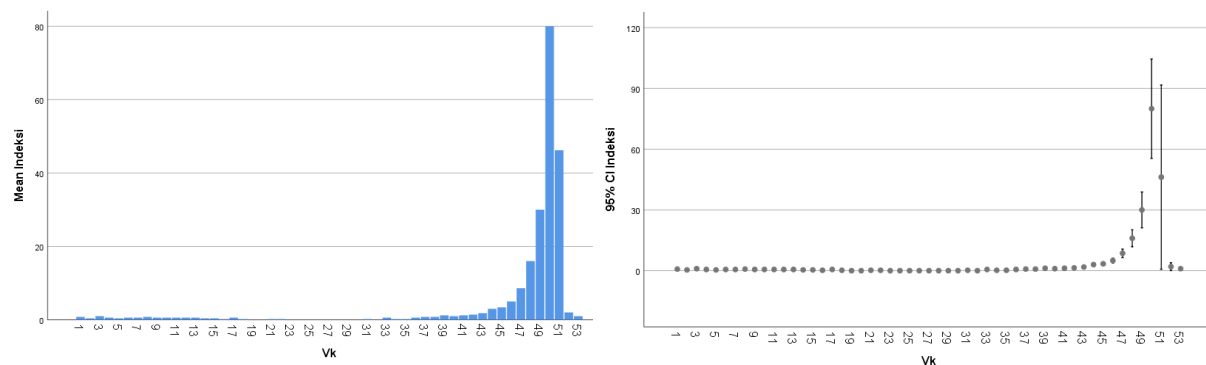


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

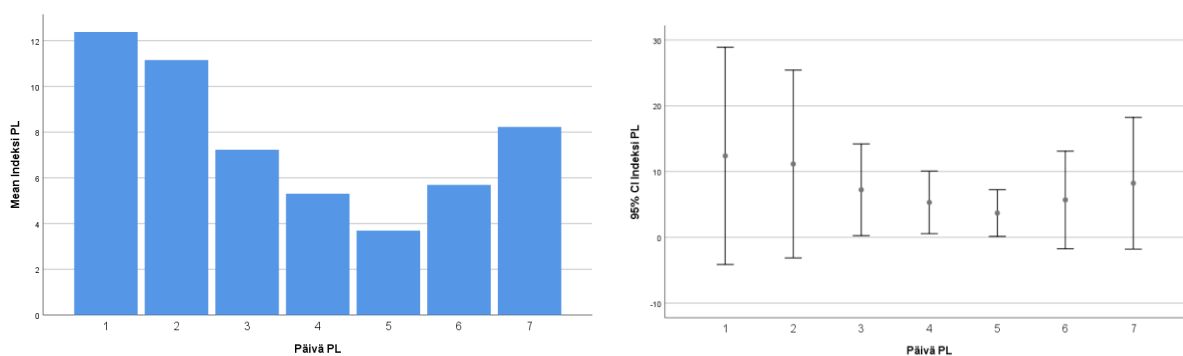
Porkkanalaatikon Google Trends -hakuindeksien viiden vuoden (2015–2019) keskiarvojen määrän jakauma kuukausittain (vasen) ja luottamusvälit (oikea)



Porkkanalaatikon Google Trends -hakuindeksien viiden vuoden (vuosien 2019–2015) keskiarvojen määrän jakauma viikoittain (vasen) ja luottamusvälit (oikea)



Porkkanalaatikon hakeminen viikonpäivittäin 90 päivän keskiarvona (Google Trends 18.2.2020) (vasen) ja luottamusvälit (oikea)

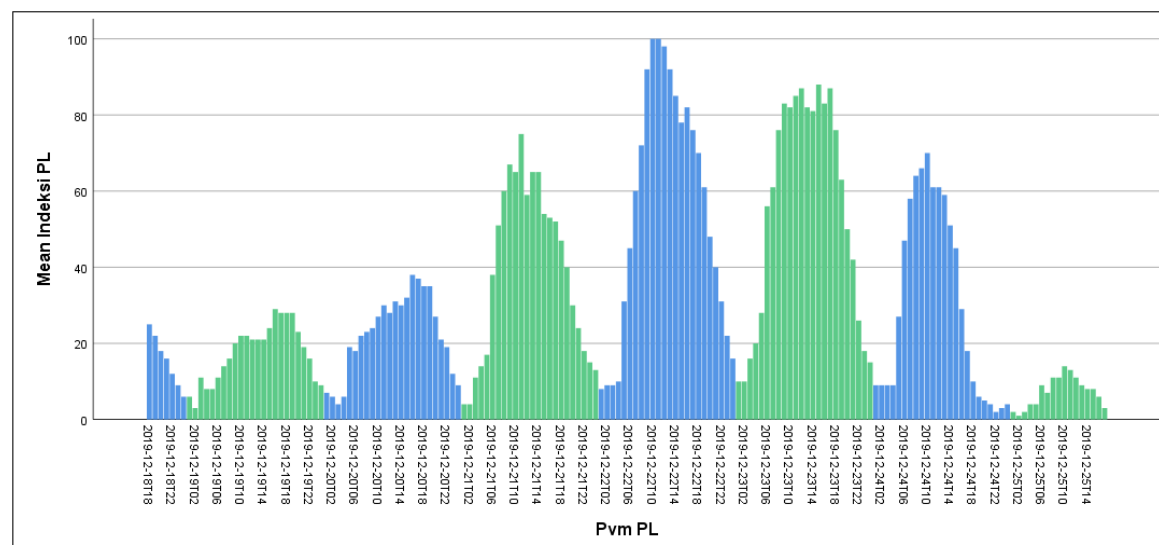


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

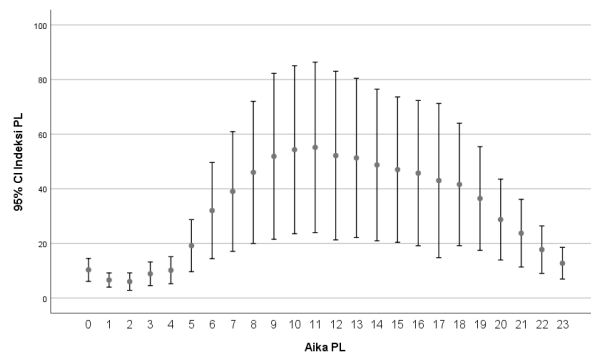
Porkkanalaatikon hakeminen viimeisen 7 päivän aikana (Google Trends 25.12.2019)



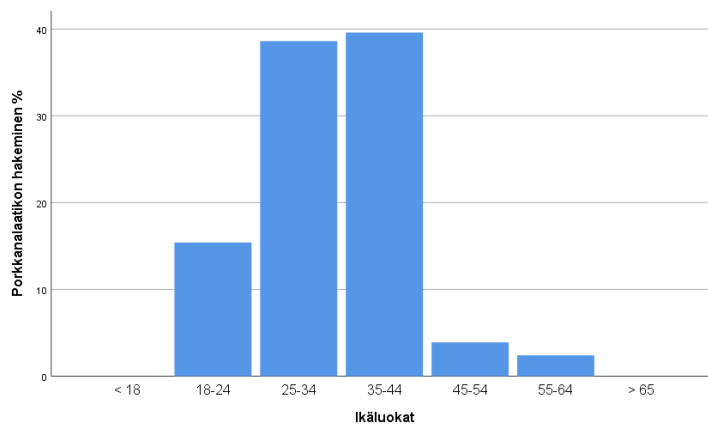
Porkkanalaatikon hakuindeksit päivittäin ja kellonajoittain joulusesongin hakuhiippuna viikon ajanjaksolla (Google Trends 25.12.2019)



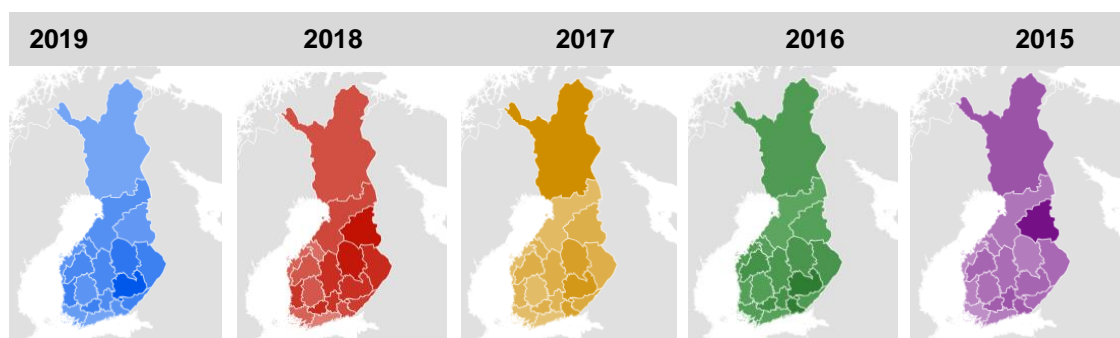
Porkkanalaatikon hakuindeksien luottamusvälit kellonajoittain joulusesongin hakuhiip-
puna viikon ajanjaksolla (Google Trends 25.12.2019)



Porkkanalaatikon hakeminen ikäluokittain (Ubersuggest 25.1.2020)



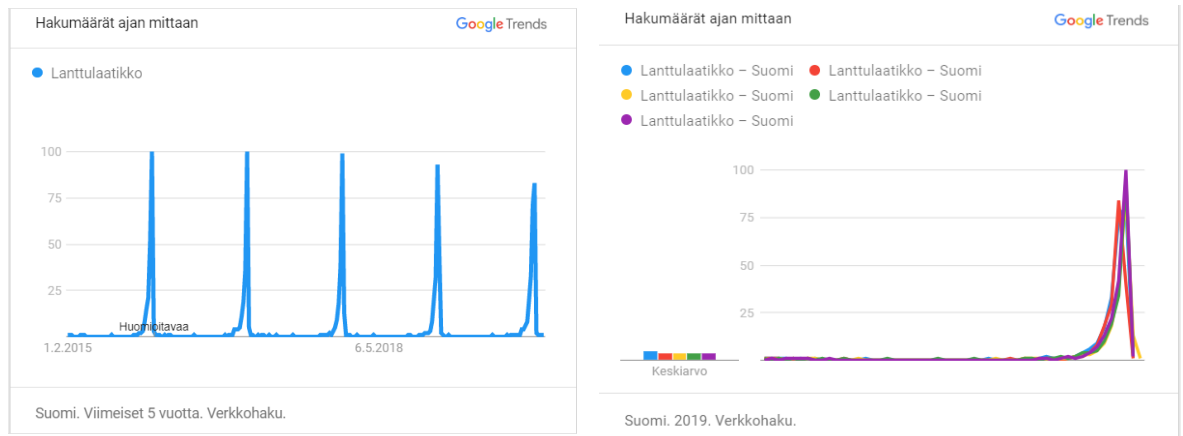
Porkkanalaatikon hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)



Keski-Pohjanmaa	
Pohjanmaa	
Etelä-Savo	
Satakunta	
Pohjois-Savo	
Kanta-Häme	
Pohjois-Pohjanmaa	
Pohjois-Karjala	
Etelä-Pohjanmaa	
Pirkanmaa	
Keski-Suomi	
Etelä-Karjala	
Päijät-Häme	
Varsinais-Suomi	
Uusimaa	
Kymenlaakso	
Kainuu	
Lappi	

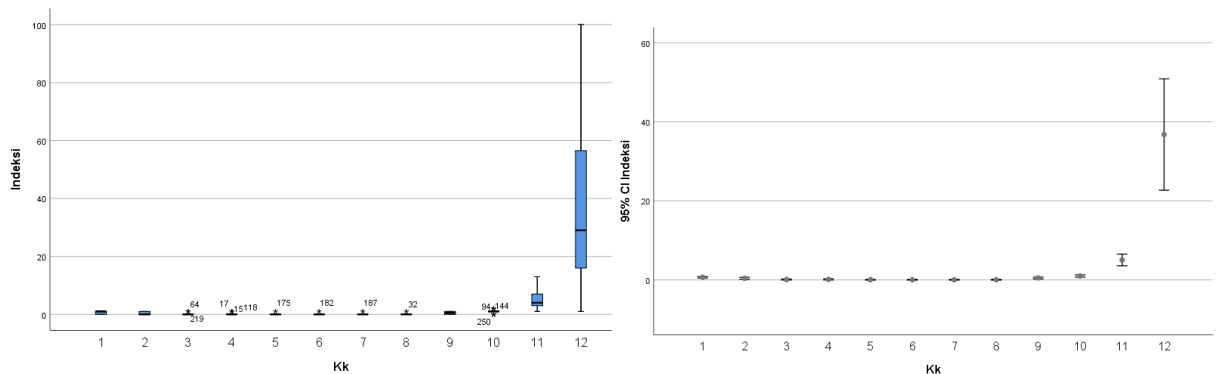
Liite 20. Lanttulaatikon hakuajankohdat

Lanttulaatikon hakemisen trendi Suomessa viiden vuoden ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

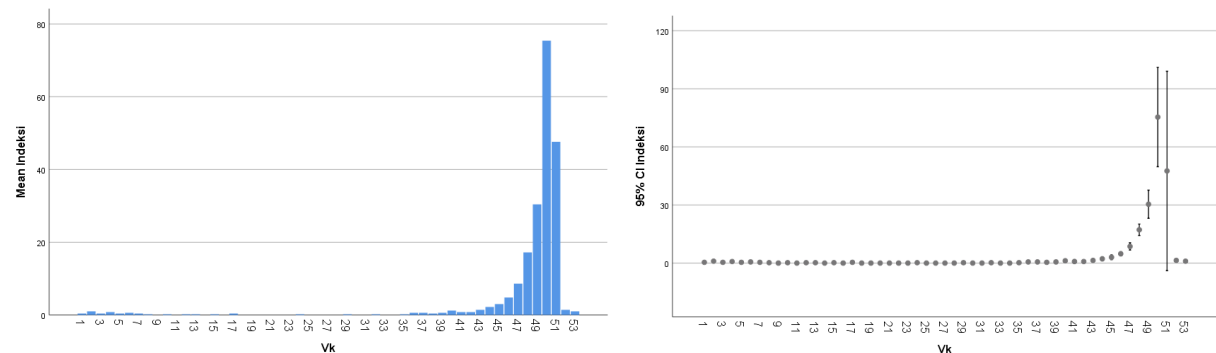


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

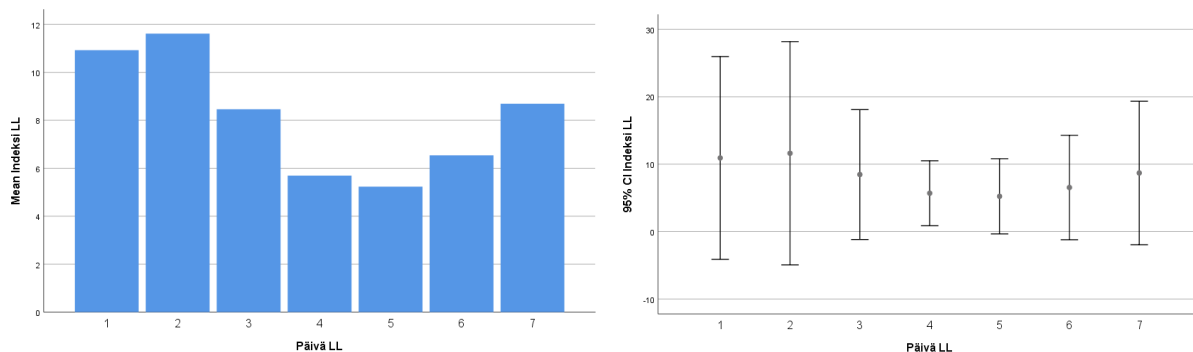
Lanttulaatikon Google Trends -hakuindeksien viiden vuoden (2015–2019) keskiarvojen määrän jakauma kuukausittain (vasen) ja luottamusvälit (oikea)



Lanttulaatikon viiden vuoden keskiarvojen indeksien määrän jakauma viikoittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Lanttulaatikon hakeminen viikonpäivittäin 90 päivän keskiarvona (Google Trends 18.2.2020) (vasen) ja luottamusvälit (oikea)

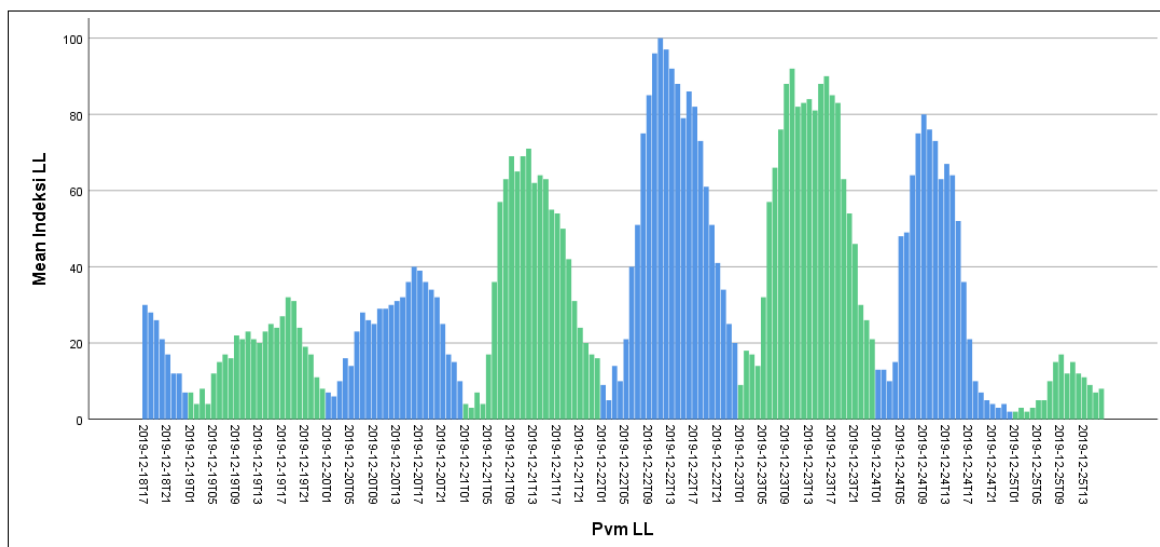


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

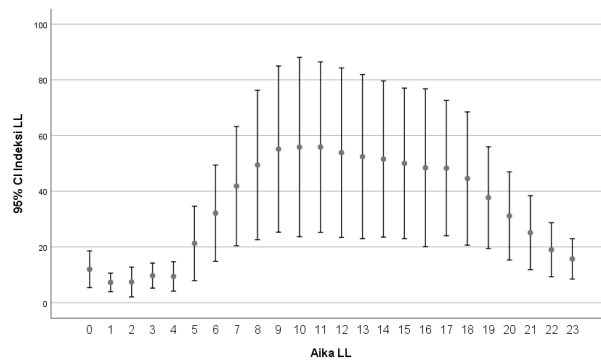
Lanttulaatikon hakeminen viimeisen 7 päivän aikana (Google Trends 25.12.2019)



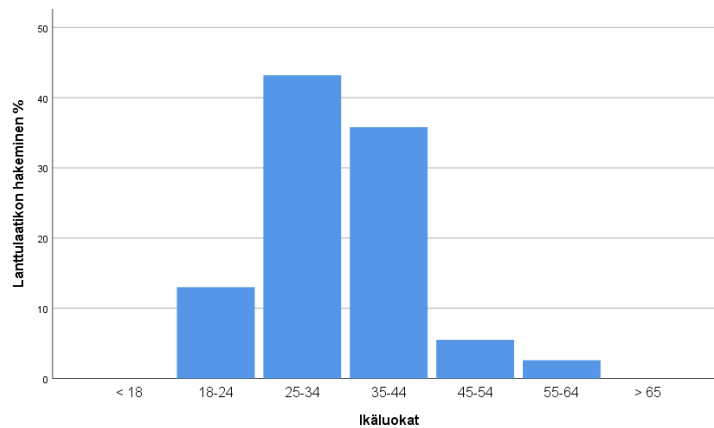
Lanttulaatikon hakuindeksit päivittäin ja kellonajoittain sesongin hakuhiippuna viikon ajanjaksolla (Google Trends 25.12.2019)



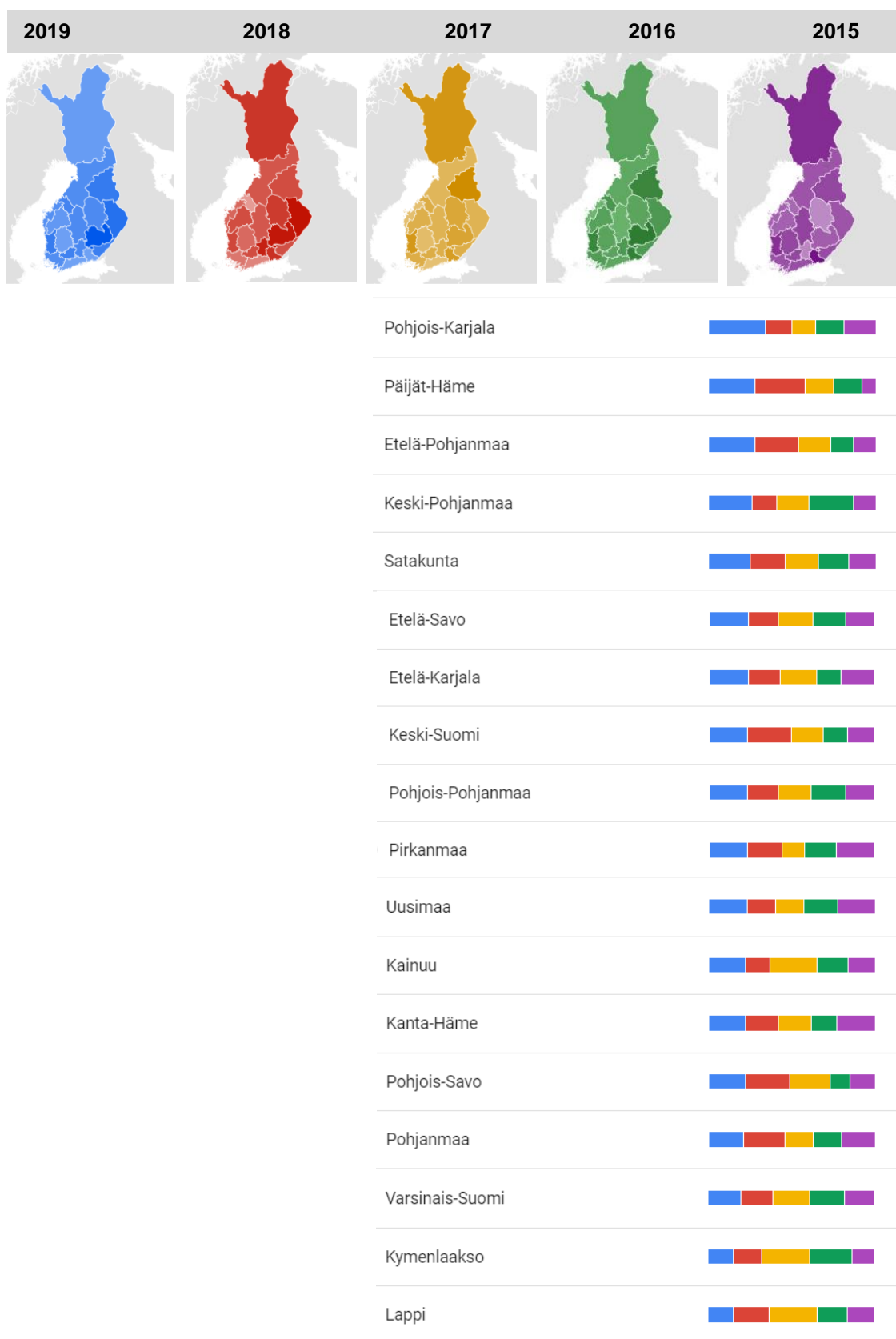
Lanttulaatikon hakuindeksien kellonaikojen luottamusvälit sesongin hakuhiippuna viikon ajanjaksolla (Google Trends 25.12.2019)



Lanttulaatikon hakemisen ikäluokat (Ubersuggest 25.1.2020)

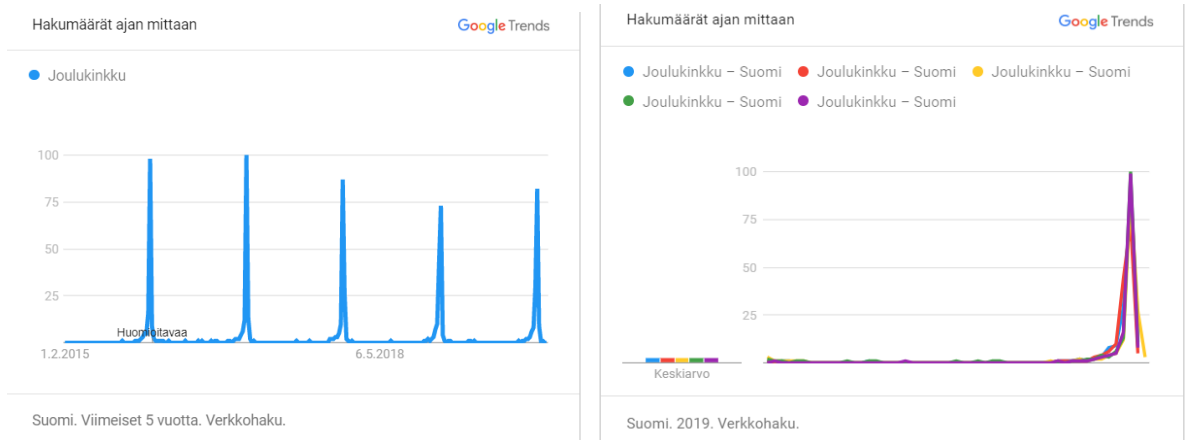


Lanttulaatikon hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)



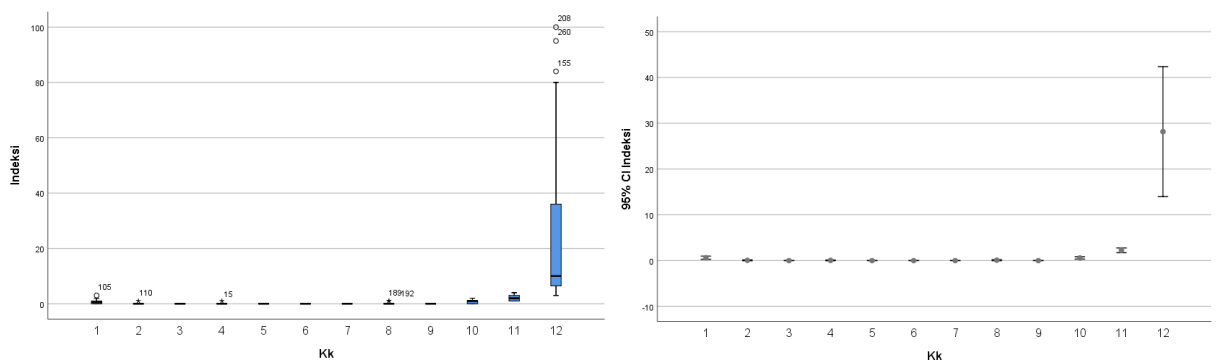
Liite 21. Joulukinkun hakuajankohdat

Joulukinkun Google Trends -haut viiden vuoden ajalla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

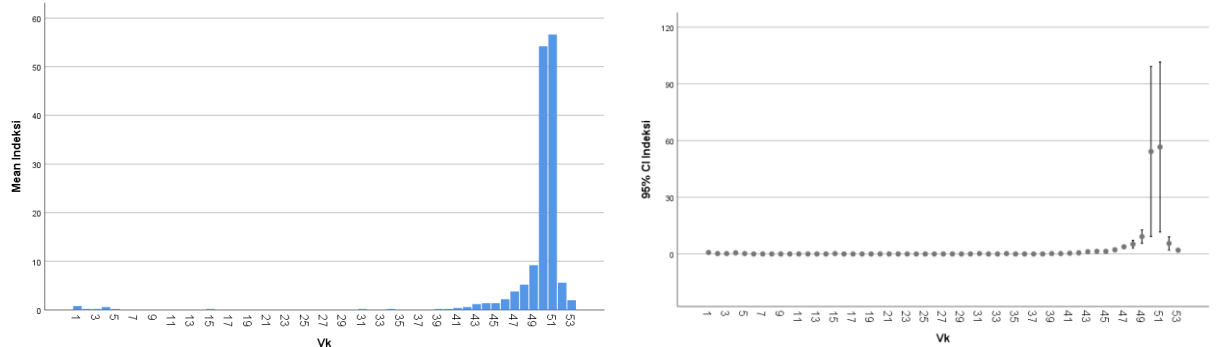


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

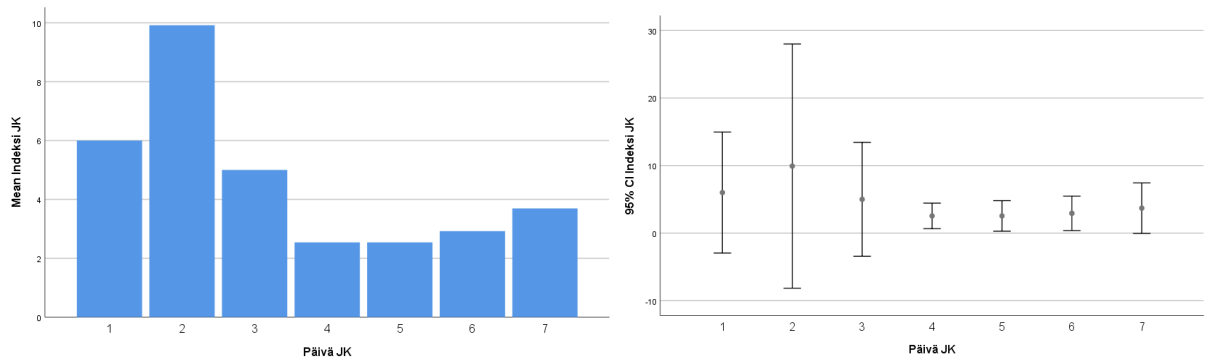
Joulukinkun Google Trends -hakuindeksien 5 vuoden (2015–2016) keskiarvojen indeksien määrän jakauma kuukausittain (vasen) ja luottamusvälit (oikea)



Joulukinkun 5 vuoden keskiarvojen indeksien määrän jakauma viikoittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Joulukinkun hakeminen viikonpäivittäin 90 päivän keskiarvona (Google Trends 18.2.2020)
(vasen) ja luottamusvälit (oikea)

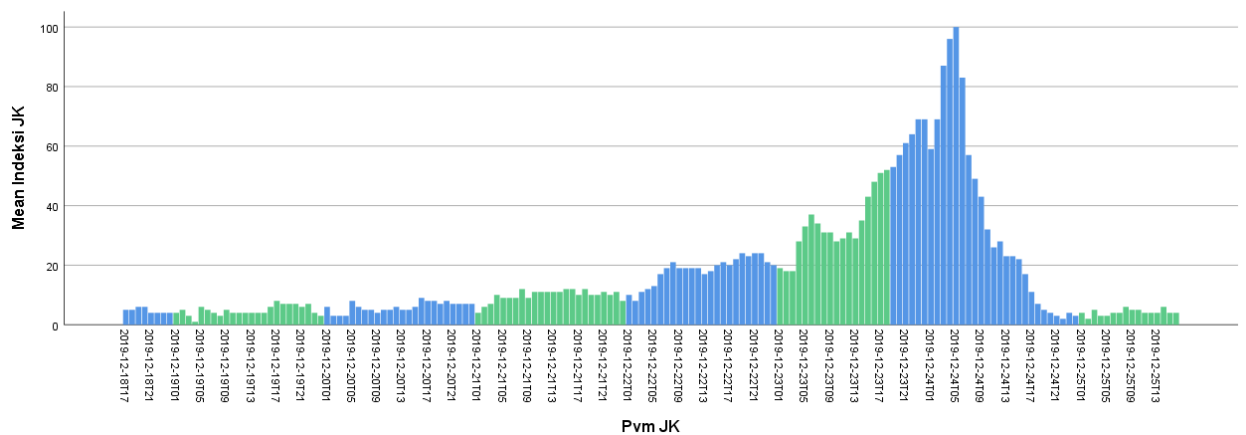


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

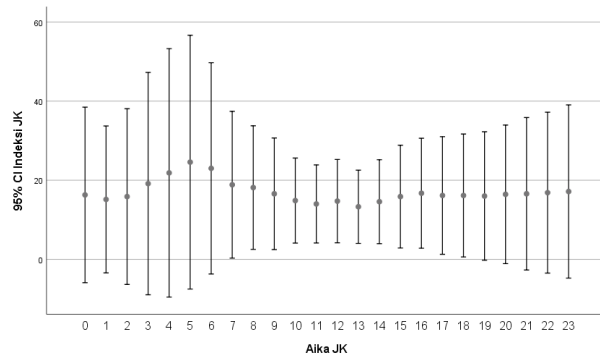
Joulukinkun hakeminen viimeisen 7 päivän aikana (Google Trends 22.1.2010)



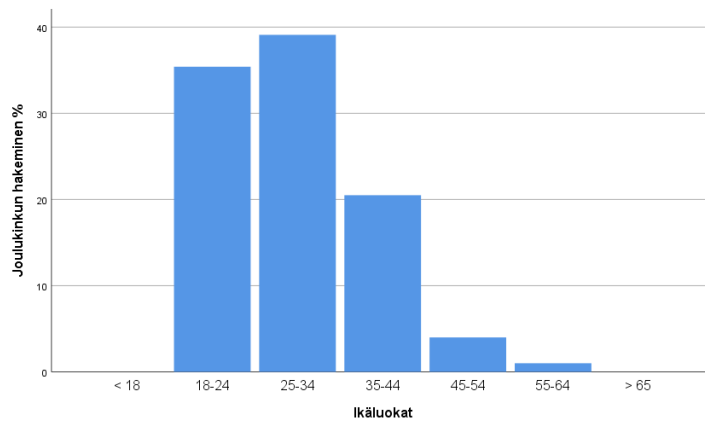
Kuvio x. Joulukinkun hakuindeksit päivittäin ja kellonajoittain seitsemän päivän aikana joulusesongin hakuhiippuna (Google Trends 25.12.2019)



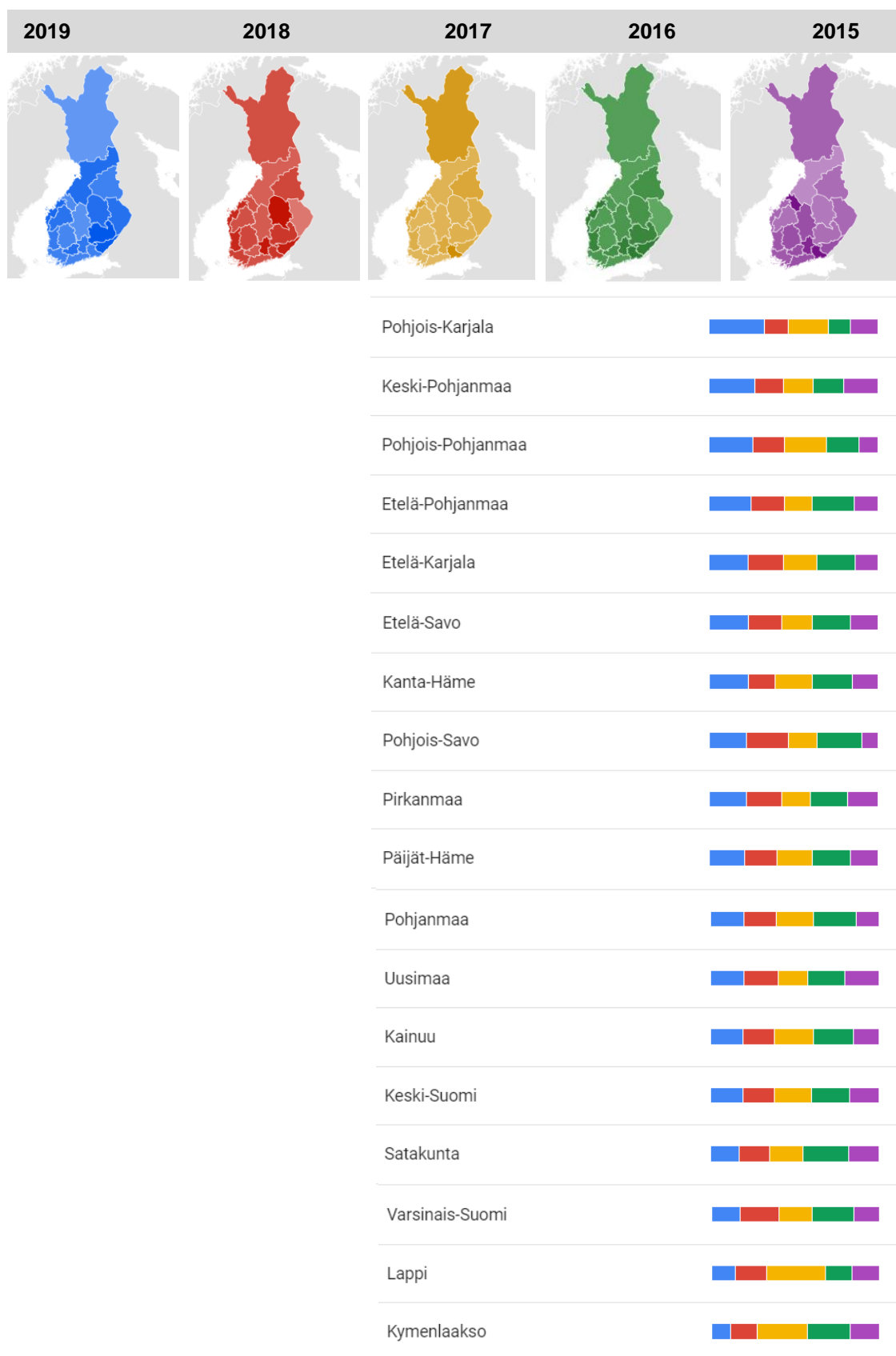
Joulukinkun hakuindeksien kellonaikojen luottamusvälit seitsemän päivän aikana jouluse-
songin hakuhuippuna (Google Trends 25.12.2019)



Joulukinkun hakeminen ikäluokittain (Ubersuggest 25.1.2020)

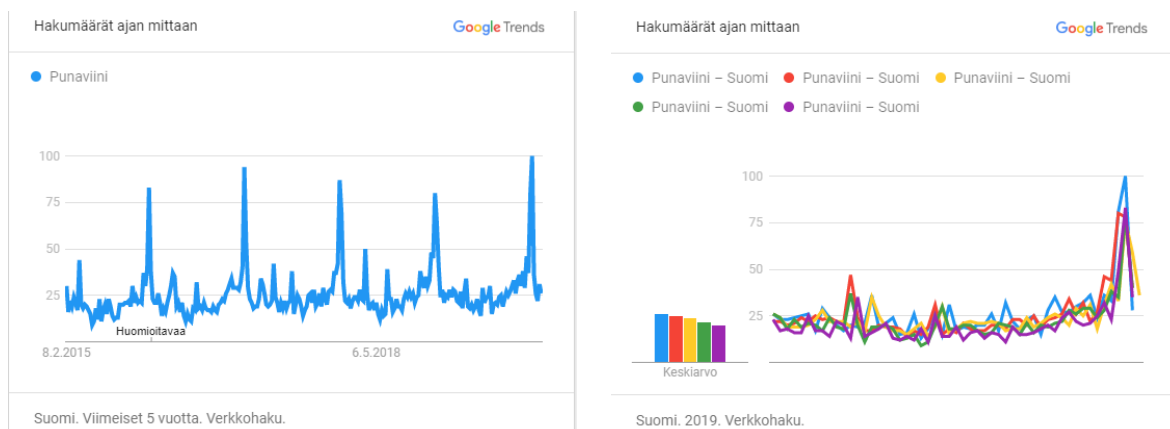


Joulukinkun hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.12.2020)



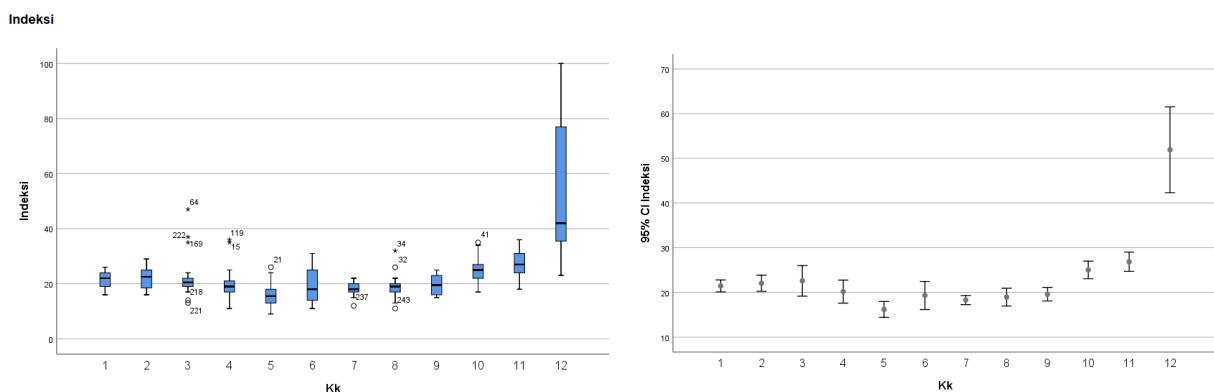
Liite 22. Punaviinin hakuajankohdat

Punaviinin hakemisen trendi Suomessa viiden vuoden (2015–2019) ajanjaksolla (Google Trends 1.2.2020) (vasen) ja vuosina 2019–2015 (Google Trends 13.2.2020) (oikea)

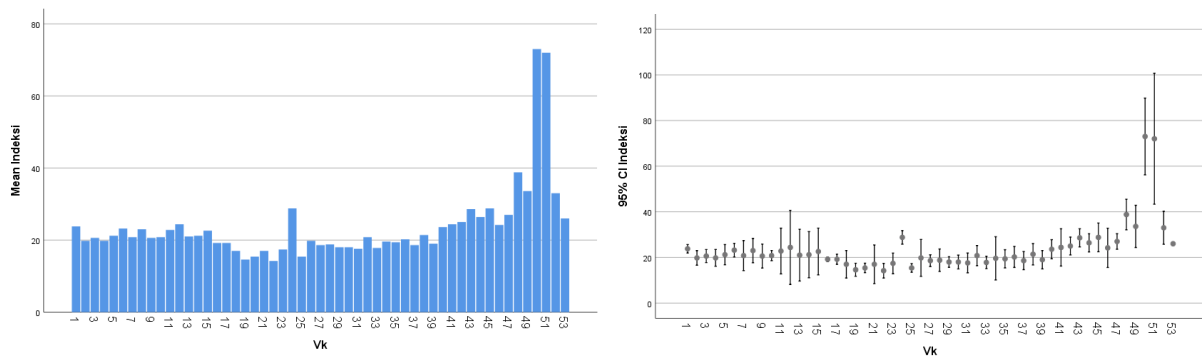


Oikeanpuoleisen kuvion värien selitykset: vuosi 2019 sininen, vuosi 2018 punainen, vuosi 2017 keltainen, vuosi 2016 vihreä, vuosi 2015 violetti

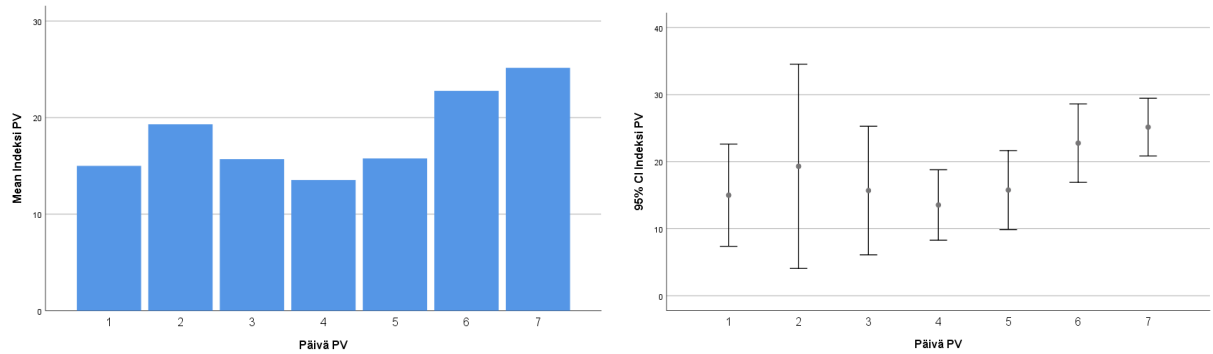
Punaviinin Google Trends -hakuindeksien viiden vuoden (2015–2019) keskiarvojen jakauma kuukausittain (Google Trends) (vasen) ja luottamusvälit (oikea)



Punaviinin 5 vuoden indeksien keskiarvojen jakauma viikonnumeroittain (Google Trends 1.2.2020) (vasen) ja luottamusvälit (oikea)



Punaviinin hakeminen viikonpäivittäin edellisen 90 päivän indeksien keskiarvona (Google Trends 22.1.2020) (vasen) ja luottamusvälit (oikea)

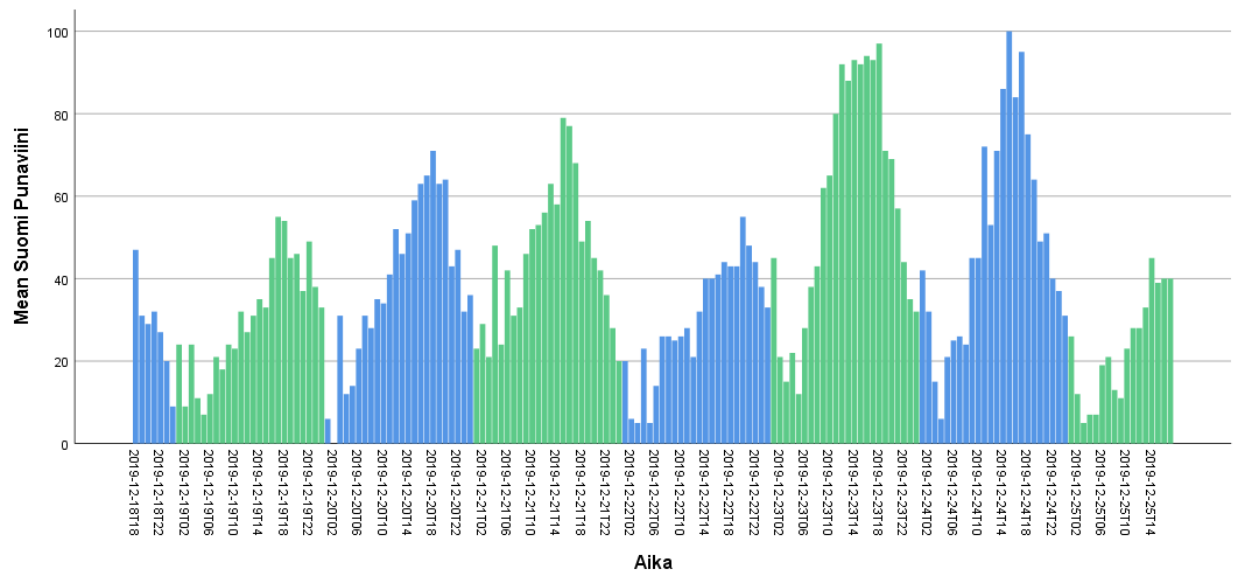


X-akselin numeroiden selitykset: 1 sunnuntai, 2 maanantai, 3 tiistai, 4 keskiviikko, 5 torstai, 6 perjantai, 7 lauantai

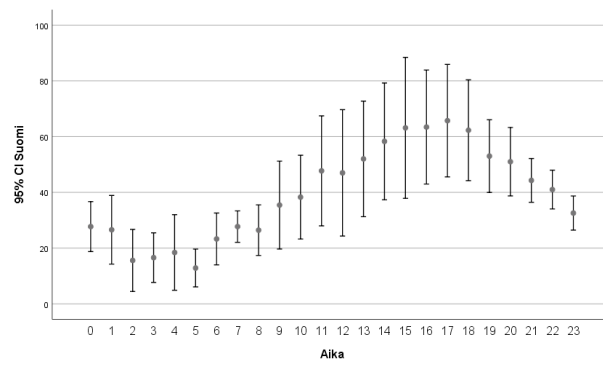
Punaviinin hakeminen seitsemän päivän aikana (Google Trends 25.12.2019)



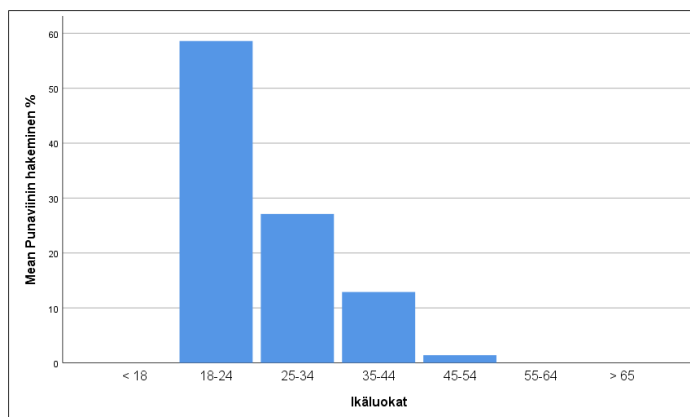
Punaviinin hakuindeksit päivittäin ja kellonajoittain seitsemän päivän aikana joulusesongin hakuhiippuna (Google Trends 25.12.2019)



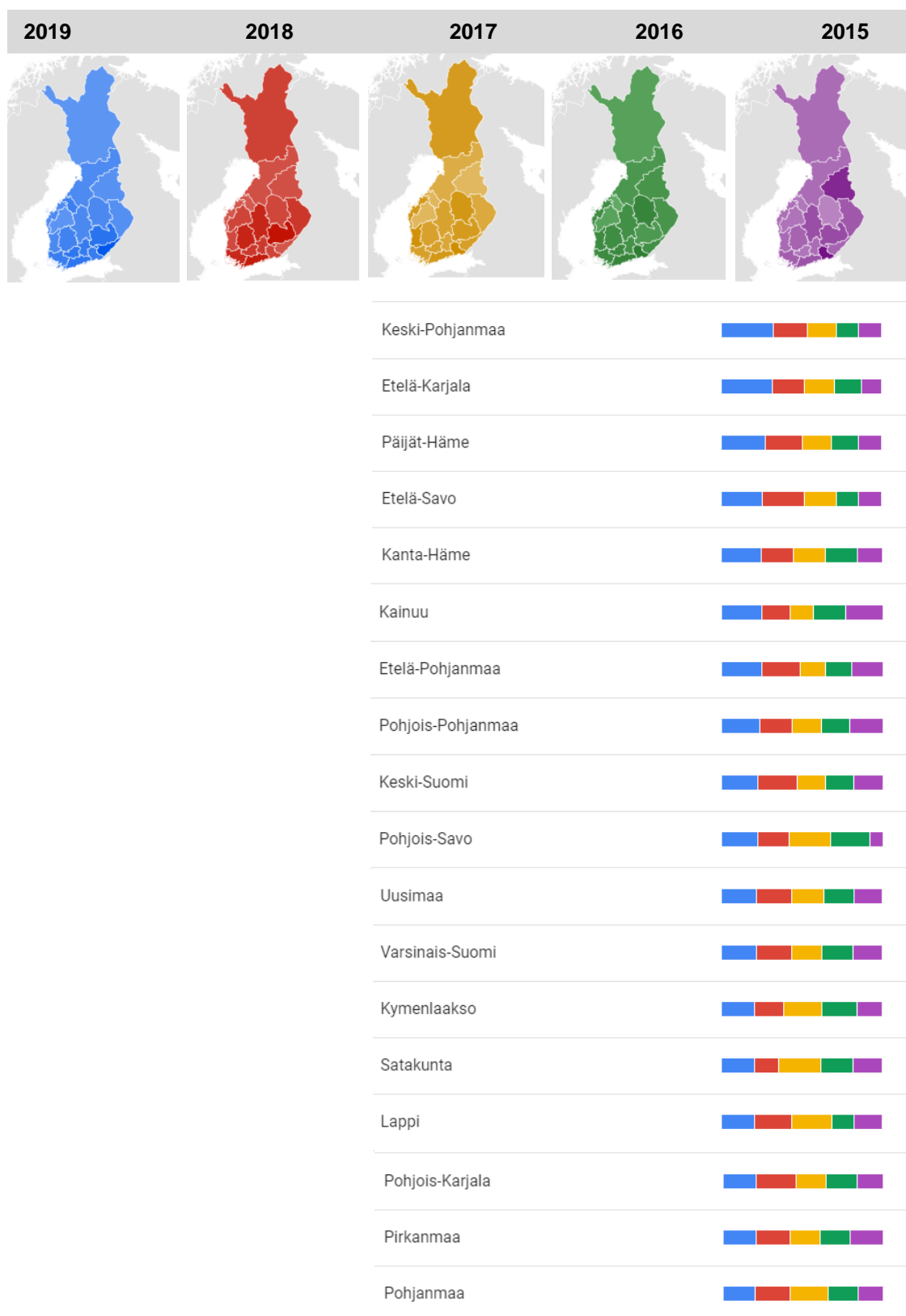
Punaviinin hakuindeksien kellonaikojen luottamusvälit seitsemän päivän aikana jouluse-
songin hakuhuippuna (Google Trends 25.12.2019)



Punaviinin hakemisen ikäluokat (Ubersuggest 25.1.2020)



Punaviinin hakemisen aluekohtaiset erot eri vuosina (Google Trends 14.2.2020)



Salaiset liitteet

Salainen liite 1. Kehittämiprojektin suunnitelma

Salainen liite 2. Kohdeyrityksen kehitysjohtajan Anon haastattelulomake

